Sasson, Amir (Ed.)

Book

At the forefront, looking ahead : research-based answers to contemporary uncertainties of management

Reference: (2018). At the forefront, looking ahead : research-based answers to contemporary uncertainties of management. Oslo : Universitetsforlaget. doi:10.18261/978-82-15-03158-3-2018.

Open access

Amir Sasson (Ed.)

At the Forefront, Looking Ahead

Research-Based Answers to Contemporary Uncertainties of Management

1943 BI 75 2018

UNIVERSITETSFORLAGET

At the Forefront, Looking Ahead

Amir Sasson (Ed.)

At the Forefront, Looking Ahead

Research-Based Answers to Contemporary Uncertainties of Management

Universitetsforlaget

This book was first published in 2018 by Universitetsforlaget.

This book is published with the support of BI Norwegian Business School.

ISBN printed edition (print on demand): 978-82-15-03140-8
ISBN electronic pdf-edition: 978-82-15-03158-3

DOI: 10.18261/978-82-15-03158-3-2018

Enquiries about this publication may be directed to:
post@universitetsforlaget.no
www.universitetsforlaget.no

Cover: BI/Universitetsforlaget
Prepress: Laboremus Sandefjord AS

Contents

DOI: 10.18261/9788215031583-2018-01

Preface

BI Norwegian Business School is an independent, not-for-profit foundation and the main provider of research-based knowledge on business and management disciplines in Norway. Since the school's establishment in 1943, BI has grown to become one of Europe's largest business schools, with around 20,000 students and 900 employees. BI is the largest supplier of economic and administrative competences and skills in Norway, with more than 200,000 graduates since 1983. BI Norwegian Business School is the home of world-renowned experts in the school's areas of specialization, including finance, economics, management, strategy, and marketing.

BI has undergone tremendous change over the last few decades. Having started as an evening school for professionals, it has become one of Europe's most advanced research institutions, attracting world-class researchers, students, and speakers, while educating professionals to lead society forward. This book is a testimony to this transformation. Its publication corresponds with BI Norwegian Business School's 75th anniversary celebrations, which are focused on our devotion to research and knowledge creation for the benefit of society at large.

Oslo, August 2018
Amir Sasson

DOI: 10.18261/9788215031583-2018-02

Chapter 1
Introduction

AMIR SASSON

EXCITING AND UNCERTAIN TIMES

Arguably, we live in the most exciting and alarming period of transformation in human history. Within less than two generations, ever since Ray Tomlinson sent the first email in 1971—or three generations if we want to start the clock at the invention of the first digital computer—information technology has penetrated almost every aspect of our lives. Unlike the industrial revolution that primarily benefited richer nations and privileged individuals, diffused through colonial powerhouses, 70% of the world's youth are online and 58% of the entire world population has a mobile broadband subscription (International Telecommunication Union, 2017). The current technological revolution transforms how we search for information, learn (from Wikipedia to online courses), spend our spare time (e.g., Candy Crush, Pokémon GO, YouTube, Facebook), produce (from molding to additive manufacturing), what we learn (limited textbooks to the Internet), and with whom we socialize (e.g., Snapchat, Tinder), to name merely a few aspects.

Business is no less affected. "Information is king" has dethroned "cash is king." Cash may still retain a certain royal status, but the crown sits serenely on the heads of those who control information. Traditional cash-rich industries (e.g., media, photography) have experienced the rapid depletion of positive cash flows when the fundamental economic principles of their cash-generating business models fell apart (Afuah & Tucci, 2003; Amit & Zott, 2012; Kolbjørnsrud, Amico, & Thomas, 2016). In other industries, traditional businesses are feeling the earth trembling beneath their feet. It is not that such traditional economic activities are necessarily disappearing; it is merely that production and service provisioning are becoming secondary in the new balance of power, whereby those who possess information also substantially influence value appropriation.

New business models primarily based on mediating technology (Sasson, 2008; Stabell & Fjeldstad, 1998; Thompson, 1967) have taken over as the economic superpowers. In 2018, Apple Inc., Alphabet Inc., Microsoft, Amazon, and Tencent have the highest market capitalizations in the world. Seven of the largest 10 publicly traded firms are a direct creation of the current technological revolution. In the

1970s, 1980s and 1990s, the largest firms were oil producers, industrial conglomerates and retailers (with the exception of IBM and AT&T, both dethroned long ago).

It should be noted that the center of gravity in terms of the underlying economic knowledge of such firms and markets has been at UC Berkeley (e.g, Katz & Shapiro, 1985; Shapiro & Varian, 1998), NYU (Economides, 1996) and at BI Norwegian Business School (Fjeldstad & Andersen, 2003; Fjeldstad, Snow, Miles, & Lettl, 2012; Stabell & Fjeldstad, 1998). At UC Berkeley, network economics developed hand-in-hand with the most vibrant information technology cluster in the world (Saxenian, 1994). The existence of world-class universities developing cutting-edge technologies, coupled with dedicated researchers in locally-present business schools, has created positive reinforcing loops further increasing cluster attractiveness (Reve & Sasson, 2012). Such symbiotic relationships have had a tremendous effect on the education of new generations of mediating firms' managers, public policy, court rulings and legislation that at the time lacked a fundamental understanding of network economics. At BI Norwegian Business School, the understanding of network business models developed in spite of the lack of a sizeable local industry and world-leading cluster in the vicinity.

When technological developments dramatically influence individual behavior on the one hand, and markets on the other, the daily life of corporations are not immune to the upheaval. Let us merely examine some examples from finance. Traditional financing arrangements are complemented and threatened by crowdfunding, blockchain and various corporate and non-corporate venture capital firms (Dushnitsky & Shapira, 2010). Investment, which is commonly home biased (Coval & Moskowitz, 1999), is now facing not only information overload but also large sovereign-backed investors that control roughly 10% of globally traded stocks and who are prohibited from investing at home (See also: Capapé & Santiso, 2017).

We will discuss additional matters that are pressing for corporations below. First, let us just admit the obvious: We are bewildered, overwhelmed, we fear redundancy and are largely in the dark on how to succeed in this new era. It does not become any easier when some argue that we are currently observing the last generation of the simple, non-upgraded, humans (Harari, 2016), that traditional long-tail manufacturing will be replaced by instantaneously automated production (Anderson, 2012; Sasson & Johnson, 2016), and that artificial intelligence will either transform or render obsolete a wide range of work tasks and management practices (Kolbjørnsrud et al., 2016).

Corporations are soon to start competing on employing the millennium generation who are digital natives, but the former know very little about how to motivate

and integrate the latter. Corporations know even less about how to convert the current non-digital-natives into digitally capable employees. Employees have misconceptions about what it takes to excel (Hansen, 2018) and which tools will assist them in doing so.

Public policy is, as usual, lagging behind. Knowledge-based economic policies are creeping slowly into contemporary economic life. Very few countries acknowledge the public aspect of life-long learning, hence creating a situation whereby employers have no incentive to invest in mobile employees who themselves are not allocated the time to invest in knowledge upgrades. Furthermore, school curricula have remained largely unchanged over the last generation, exposing the millennium generation to largely redundant languages instead of the language of tomorrow: data.

We live in an era of the unprecedented availability of information, yet we have only just started learning how to create and appropriate value (Magretta, 2002) from information. In the era of the explosion and distortion of information, we need islands of sanity: dedicated researchers who devote their time and effort to providing reliable and valid evidence and thought-provoking ideas about how to understand, create and adapt to these unparalleled life-changing developments.

This book presents research-based answers—not *the* answers—to some of the uncertainties that managers, investors, employees and policy makers face in this new era. As the title, "At the Forefront, Looking Ahead" indicates, on its 75th anniversary and having started as a practitioner-oriented evening school, BI Norwegian Business School is undoubtedly a research-based school at the forefront of global research. Its research groups contribute excellent, original research that is at the international forefront and appears in outstanding international journals (The Research Council of Norway, 2018), while its graduates, more than any other school, populate CEO positions in the largest 500 local firms.

Being at the forefront requires that we look ahead, not merely celebrate past successes. The book does exactly that. It covers three themes: 1) The D-G-tal organization, including algorithm-based decision making and management, digital labor, business models, corporate reputation and branding; 2) The governance of corporations, with specific reference to state-owned and family-owned firms and their auditing; 3) Decision making, incentives and innovation, covering issues such as employee motivation and creativity, environmental R&D, political decision making and customer experience.

The first section, the D-G-tal Organization, commences with the case for arguably the greatest transformation of organizations and managers. In chapter 2, Andersen, Johnson, Kolbjørnsrud and Sannes explore the managerial, organiza-

tional, and strategic implications of letting algorithms and learning systems based on massive amounts of data make increasing number of organizational decisions. These "intelligent enterprises," with enhanced abilities to sense, comprehend, act, learn and explain (SCALE) the environment, must cede authority over some decisions, while developing SCALE capabilities and roles to stay competitive.

Wong and Fieseler discuss the meaning and implication of digital work. Chapter 3 discusses how digital technologies transform contemporary labor arrangements. In particular, they identify a) aspirational labor, whereby future employees exhibit work-related skills providing a track record towards building a digital CV; b) platform labor, whereby platforms are mediated between individuals who maintain their autonomy and offer skills to contract providers, resulting in highly skilled, temporary and mobile labor and; c) corporate labor, which introduces arrangements aimed at increasing efficiency, such as computer-mediated communication tools and enterprise social media.

That reputation is intangible, inimitable, and difficult to acquire means that it has the qualities to be a valuable resource in supporting a firm's quest towards competitive advantage. It also has characteristics that make it very fragile. While it takes a prolonged period to establish a positive reputation, when abundant information can be instantaneously diffused, it may only take minutes or hours to shatter it. In Chapter 4, Brønn and Buhmann offer insights into the challenges of building, maintaining and managing organizational reputation.

Branding has not been left untouched by digitalization. In Chapter 5, Olsen simply asks the million-dollar question: To what extent are insights from the traditional strategic understanding of branding still relevant in the digital age? Olsen argues that new digital technologies, media channels and online consumption patterns provide a realm of new opportunities to brand managers. Brand managers should continue developing a deep understanding of what consumers actually need, not just which data brand managers can gain in the short run.

Digital transforms markets, organizations, and organizing. Fjeldstad and Haanæs address a vital issue for firms: how value creation and organization design are affected by digitalization. Chapter 6 takes us on a journey through various business models and discusses how digital may radically transform the mechanisms by which activities and resources are differentiated and integrated for each business model type. They also report on the new organization design principles for organizing digitally.

The second section explores corporate governance. In Chapter 7, Berzins, Bøhren and Stacescu report on empirical evidence of the most common and most neglected form of organizing, namely the family firm. Taking an agency theory

perspective, the authors argue that two common agency problems are, to a large extent, non-existent in family firms. When firm owners and insiders are one and the same, the first agency problem—conflict of interests between these two groups—does not arise. When 82% of family firms have no minority owners, the second agency problem—whereby the majority can act in its interest at the expense of the minority—does not exist, and it is of lesser magnitude when the owners of many of the remaining family firms control stakes exceeding two-thirds. This may be an explanation for the persistent finding of higher performance in family firms compared to non-family firms.

Chapters 8 and 9 address legal and historical aspects of state ownership. The State is the largest owner on the Norwegian Stock Exchange. Christensen traces the development of the Norwegian state ownership model to the State's ownership of Norsk Hydro. The author argues that the model, which entails that the state act as a private shareholder—respecting minority shareholder rights and taking a more long-term strategic view—explicates the wide-ranging support for the state as a legitimate owner. Bråthen complements the above by arguing that one of the reasons for the success of state ownership of publicly listed firms is the gradual development of the legal framework for the exercise of state ownership. Having said that, Bråthen identifies some unresolved legal aspects of state ownership, including active ownership, the status of the State's principles for corporate governance, and potential liability of the State for damages.

A related topic is the functioning of the auditing system, which mitigates agency conflicts between firm directors who produce financial statements, and investors, such as the State, that use them. Regulation is necessary to ensure the independence of auditors who are paid by the firm and to ensure adequate quality in a concentrated market. In the light of contemporary regulatory changes, in Chapter 10 Langli and Willekens review novel regulations and highlight obstacles to the development of our understanding of the impact of auditor regulation on auditors' behavior and quality. Surprisingly, in a rich data environment such as auditing, most of auditors' work is confidential, thus preventing further development of our knowledge of auditing work and its quality. In an era characterized by distorted information, the role of external financial statement auditors in assuring reliability increases in importance.

The third section explores decision making, incentivizing and innovation. Škerlavaj argues that creativity and innovation, which are the driving forces behind the data-driven revolution (Frank, Roehrig, & Pring, 2017), are also the solution to the many societal ambiguities and uncertainties that we face. However, high-potential creative ideas are at high risk of dismissal and are often replaced by moderately

novel ideas. The author of Chapter 11 takes a leadership perspective to understand the fate of many high-potential creative ideas. Great leaders who wish to create the conditions under which high-potential creative ideas can flourish should act as change agents for innovation, integrators across units, disciplines and perspectives, should be helpful and supportive, and encourage proactive behavior.

How to motivate employees to increase productivity has been a heavily debated topic. In light of the emerging new digital workforce, the understanding of the fundamental antecedents to motivation in general, and to intrinsic motivation in particular, is of vital importance. Kuvaas reports a thought-provoking meta-study that shows that ability and motivation are similarly important to job performance, and that high levels of motivation can compensate for lower levels of ability (Van Iddekinge, Aguinis, Mackey, & DeOrtentiis, 2018). Chapter 12 provides ample of empirical evidence establishing the importance of intrinsic motivation and guides leaders through factors that affect intrinsic motivation.

Bjertnæs, Heggedal and Jacobsen discuss a contested and not fully understood R&D policy in the age of sustainability. Environmentally friendly R&D is a necessary condition in addressing environmental challenges. The current model for climate R&D policy argues for gradually declining subsidies to address the inefficiencies in this research market. In Chapter 13, the authors argue that in the presence of increasing returns to scale—for example, knowledge spillovers—increasing R&D subsidies are optimal.

In the light of recent technological changes, the understanding of how agents exert influence by strategically transmitting private information to policy makers will only gain in importance in the years to come. Actors will attempt to exert power through lobbying decision makers who do not yet know mean to end relation of information policies. Taking a game theoretical approach in Chapter 14, Helland, Monkerud and Løyning report that in a costly signaling game, elite politicians from the Norwegian National Assembly are substantially more off-mark relative to equilibrium predictions than students are. It begs the question of whether we should trust lobby-prone politicians with technology-driven transformational policies.

Finally, yet importantly, the meaning of excellent service is technology-dependent and, hence, firms' understanding of the service encounter is continuously in flux. Information technologies redefine the realm of opportunities at the hands of service providers, from in-store location tracking through electro-dermal activity, to tracking online behavior. In Chapter 15, Gustafsson and Lervik-Olsen review the evolution of service marketing and provide an intriguing integration of services and technologies.

We hope that you will enjoy reading these thought-provoking ideas and the empirical evidence that may not fit with your ex-ante presumption of business, management and decision making. Please do not hesitate to contact us with comments and suggestions[1]. Notwithstanding our position at the forefront, we are always looking toward the horizon for new opportunities for knowledge creation. The journey has just begun.

REFERENCES

Afuah, A., & Tucci, C. L. (2003). A model of the internet as creative destroyer. *IEEE Transactions on Engineering Management, 50*(4), 395–403.

Amit, R., & Zott, C. (2012). Creating value through business model innovation. *MIT Sloan Management Review, 53*(3), 41–49.

Anderson, C. (2012). *Makers: The new indsutrial revolution*. London: Random House Business Books.

Capapé, J., & Santiso, J. (2017). *Sovereign wealth funds 2017*. Retrieved from https://www.ie.edu/ie-foundation/applied-research/projects/sovereign-wealth-funds-2017/

Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance, 54*(6), 2045–2073.

Dushnitsky, G., & Shapira, Z. (2010). Entrepreneurial finance meets organizational reality: Comparing investment practices and performance of corporate and independent venture capitalists. *Strategic Management Journal, 31*(9), 990–1017.

Economides, N. (1996). The economics of networks. *International Journal of Industrial Organization, 14*(6), 673–699.

Fjeldstad, Ø. D., & Andersen, E. (2003). Casting off the chains *European Business Forum, 14*, 47–53.

Fjeldstad, Ø. D., Snow, C. C., Miles, R. E., & Lettl, C. (2012). The architecture of collaboration. *Strategic Management Journal, 33*(6), 734-750. doi:10.1002/smj.1968

Frank, M., Roehrig, P., & Pring, B. (2017). *What To Do When Machines Do Everything: How to Get Ahead in a World of AI, Algorithms, Bots, and Big Data*. John Wiley and Sons.

Hansen, M. T. (2018). *Great at Work: How Top Performers Do Less, Work Better and Achieve More*. UK: Simon & Schuster UK Ltd.

Harari, Y. N. (2016). *Homo deus: a brief history of tomorrow.* HarperCollins Publishers.

International Telecommunication Union. (2017). *ICT facts and figure 2017*. Retrieved from https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf

Katz, M. L., & Shapiro, C. (1985). Network externalities, competition and compatibility. *American Economical Review, 75*(3), 424–440.

---

1. We would like to thank the authors for their dedication and contribution, as well as the book committe members, Dagmar Langeggen and Audun Farbrot, for their irreplaceable role in initiating, forming, and finalizing this project.

Kolbjørnsrud, V., Amico, R., & Thomas, R. J. (2016). How artificial intelligence will redefine management. *Harvard Business Review*.

Magretta, J. (2002). Why business models matter. *Harvard Business Review, 80*(5), 86–92.

Reve, T., & Sasson, A. (2012). *Knowledge-based Norway (original title: Et Kunnskapsbasert Norge)*. Oslo, Norway: Universitetsforlaget.

Sasson, A. (2008). Exploring mediators: Effects of the composition of organizational affiliation on organization survival and mediator performance. *Organization Science, 19*(6), 891–906. doi:10.1287/orsc.1080.0363

Sasson, A., & Johnson, J. C. (2016). The 3D printing order: variability, supercenters and supply chain reconfigurations. *International Journal of Physical Distribution & Logistics Management, 46*(1), 82–94.

Saxenian, A. (1994). *Regional advantage: Culture amnd competition in Silicon Valley and Route 128.* Cambridge, MA: Harvard University Press.

Shapiro, C., & Varian, H. R. (1998). Versioning: The smart way to sell information. *Harvard Business Review, 76*(6), 107–114.

Stabell, C. B., & Fjeldstad, Ø. D. (1998). Configuring value for competitive advantage: on chains, shops, and networks. *Strategic Management Journal, 19*(5), 413–437.

The Research Council of Norway. (2018). *Evaluation of the Social Sciences in Norway*. Retrieved from Oslo, Norway.

Thompson, J. D. (1967). *Organizations in Action*. New York: McGraw-Hill.

Van Iddekinge, C. H., Aguinis, H., Mackey, J. D., & DeOrtentiis, P. S. (2018). A meta-analysis of the interactive, additive, and relative effects of cognitive ability and motivation on performance. *Journal of Management, 44*(1), 249–279. doi:10.1177/0149206317702220

Part I

The D-G-tal Organization

DOI: 10.18261/9788215031583-2018-03

Chapter 2
The data-driven organization: Intelligence at SCALE

ESPEN ANDERSEN, JOHN CHANDLER JOHNSON, VEGARD KOLBJØRNSRUD AND RAGNVALD SANNES

ABSTRACT Evolving at an unprecedented pace, digital technologies promise to automate not only labor-intensive and repetitive work, but also the traditional and exclusive domain of educated humans—knowledge work. This is evident in the new ways of reaching customers and coordinating activities, as well as in the fact that companies conducting a business built on the new technologies now constitute the world's largest enterprises. The presence and evolution of these companies challenge established divisions of labor between man and machine, and almost casually redraws the boundaries between industries. Machine learning and analytics challenge the managers leading and the managerial scientists studying organizations. Everybody says they want to be data driven—but what does a company really need to do to achieve that?

This article will explore the managerial, organizational, and strategic implications of allowing an ever increasing number of organizational decisions to be taken not by managers employing intuition and common sense, but by algorithms and learning systems based on massive amounts of data derived from electronically based customer interactions. We argue that these companies can be thought of as "intelligent enterprises" with enhanced abilities to sense, comprehend, act, learn and explain (SCALE) their environment and their interactions with it. To acquire these capabilities, managers need to cede authority over some decisions while acquiring new capabilities and roles for themselves.

KEYWORDS: Analytics | Digital strategy | Intelligent systems | Artificial intelligence | Organization design | Decision making

2.1 DATA, DATA EVERYWHERE

> "Change has never happened this fast before,
> and it will never be this slow again"
>
> *– G. Wood (2009)*

Data, which used to be expensive and scarce, is now everywhere. Digital customer interfaces make customers' purchases, pre-purchase searches, and post-purchase reactions available for analysis. Sensors record what happens, transmitting the information through air and fiber optics to those who want to use it. With the exponential growth in raw data comes even faster growth in our ability to do something with it: store it, process it, and present it. Moore's law (1965) applies to semiconductors, but exponential growth is everywhere in information technology (Brynjolfsson and McAfee, 2014; Denning and Lewis, 2016)—and is hard to wrap our heads around. With the rate of current growth, we will have systems with 32 times the current capacity in ten years; in twenty years, more than 1000 times.

With growth in data and computing comes a truly breathtaking increase in analytical capability. In early 2017, two of the authors created a course in analytics—and less than one year later, we have changed core analytics technologies two times and reframed the course, since the tools available have automated many operations.

An example of this is Google's AlphaGo. In March 2016, AlphaGo played the board game Go and won 4–1 against Lee Sedol, considered to be the greatest player of the past decade (Google, 2017). Eighteen months later, the next version, AlphaGo Zero, trained itself for three days and then won 100–0 over AlphaGo (Silver et al., 2017), before going on to teach itself chess in four hours and resoundingly beating all known chess programs (Sterling, 2017). The evolution does not stop—developments in quantum computing may lead to computers that are a million or more time faster than current computers (Waters, 2018). Current organizational designs and decision-processes do not have the ability to use the future capacity of computers to collect, process, and distribute information.

The rapid evolution will require organizations to be data-driven, using new tools and techniques to analyze data in order to make intelligent decisions. We think of these organizations as information processing systems (Galbraith, 1973), structured "to create the most appropriate configuration [...] to facilitate effective collection, processing and distribution of information" (Tushman and Nadler, 1978: 615), and as displaying two types of organizational innovation (Galbraith, 2010, Fjeldstad *et.al.,* 2012). First, they have developed coordinative and collaborative capabilities that are information-intensive and automated to manage increasing complexity. Second, they can self-organize and self-reconfigure on a continuous basis.

2.2 ORGANIZATIONAL INTELLIGENCE AT SCALE

To deconstruct and describe a data-driven organization, we think of it as an intelligent system that can Sense, Comprehend, Act, Learn, and Explain—SCALE—inspired by the literature on artificial intelligence (Winston, 1984; Simon and Newell, 1958). These are capabilities that characterize intelligent actors such as humans and smart machines, as well as organized collectives of both. We think this SCALE framework will allow scholars and practitioners to better understand how organizations can turn data into sustainable competitive advantage. Emphasizing the role of data and technology, the following sections explain the SCALE framework of organizational intelligence. We use the vehicle manufacturer Tesla to illustrate each SCALE element.

2.2.1 SENSE

Sensing, a dynamic organizational capability (Teece et al.,1997), involves observing and registering the external and internal environment. Sensing technologies—sensors and the technologies that connect them—enable data collection at scale from many sources. Use of digital and digitally enabled products and services leave digital traces, which become useful data. Increased analytical capabilities turn output from conventional ERP and CRM systems from irritating noise to vital information. Third-party vendors emerge, specializing in data collection, structuring, and aggregation. Furthermore, companies such as Capital One are increasingly willing to invest substantially in acquiring data by deliberately extending their services into new markets and products, less to sell there than to acquire the data required to build new products (Davenport and Kim, 2013).

Since 2016, Tesla has embedded eight 360-degree view cameras in their new cars—some of them not in use at launch, but available in anticipation of new, as yet undefined functionality. Combined with mobile connectivity, these sensors provide Tesla with a sophisticated sensing capability well beyond what conventional car manufacturers currently have—a solid basis for the other SCALE capabilities.

2.2.2 COMPREHEND

Comprehension involves using data and observations from sensing activities to discern context, detect patterns, and make inferences. Organizations can build data-driven models of their internal and external environment, identify causal relationships, and prescribe what to do. Descriptive and predictive analytics can

enhance an organization's comprehension capabilities, as can speech, image, and video recognition technologies.

The combination of sensing and comprehension enables novel data applications such as the generation of virtual representations—digital twins—of physical objects and systems that allows organizations to monitor, diagnose, and maintain such installations remotely. Companies like General Electric, Siemens, and Rolls Royce Maritime provide digital twin solutions in construction, shipping, energy, and manufacturing. Digital twins are continuously updated and often developed before their "physical twin" in order to test a product or system before it is built. Using virtual reality goggles, architects can offer customers a guided tour of a 3D virtual representation of a planned building—as was done at a newly opened hospital in Østfold, Norway—and train staff to use the new building before it is finished. Use of cheap cell phones to 3D photograph the building (generating a "point cloud") as it was assembled let the builders track progress and discover errors while they still could be inexpensively fixed.

Tesla accumulates and analyzes the vast data from their connected vehicles to identify maintenance, safety, functionality, and performance improvements, in addition to allowing services such as the remote unlocking a car by a customer. The collected data allows the company to understand how their vehicles perform when used by real customers—as well as to respond to criticism, as demonstrated when a car journalist reported the car to have limited range, and Tesla could show that the journalist had deliberately run the battery down and neglected to charge the car (Muller, 2013).

2.2.3 ACT

Action refers to the decision-making and productive activities of an organization. Technology is currently used to automate and augment processes previously reserved exclusively for humans, as well as enabling decisions and activities that used to be impossible or unfeasible to execute. The development of physical and software robots allows automatic execution of productive activities, particularly routinized processes, but with the growth in machine intelligence it is increasingly also applied to more adaptive forms of work. Banks and financial institutions, in Norway and worldwide, are rapidly implementing Robotic Process Automation (RPA) technology to automate information-based routine processes and decisions, such as handling credit card applications. Companies use chatbots with embedded natural language recognition (comprehend) and generation (act) technology to answer standard questions and offer 24/7 service to customers. Sparebank1 SR-

bank, a regional bank headquartered in Stavanger, Norway, uses a local chatbot provided by Boost.ai that reportedly understands Norwegian dialects (Lyche, 2017). Advertisers and media agencies use programmatic advertising systems to automate ad placement in different media outlets.

The Danish cafe chain Joe & the Juice has developed a system for assessing the market potential in different geographies that managers consult on a weekly basis in deciding where to launch new outlets. For instance, the chain opened 14 cafes in New York in 2017 and estimates that it will saturate that specific market with 86 outlets. Joe & the Juice uses the data from each new outlet to update the company's model, which guides where to open the next cafe in the city. The chain currently has 230 outlets in 15 countries (Andersen, 2018). Organizations personalize digital services based on user behavior and characteristics, and implement product improvement via software upgrades on existing hardware.

Acting technology can be physical, as with 3D printing used to create physical replicas of digital originals (Sasson and Johnson, 2016), i.e. spare parts for time-critical industrial equipment, hearing aids tailored to each user's ears, medical implants, prescription lenses, and even vehicles and buildings—disrupting traditional production and supply chains in the process.

All cars produced by Tesla are permanently connected to the Internet, allowing the company to update the software remotely. This service is provided free of charge to the customer and can involve some attractive freebies—free music streaming from Spotify, for instance. The effect is that customers eagerly look forward to software upgrades—and that Tesla can fix errors quickly and gain the benefit of all their cars running on the latest software, drastically reducing model and version complexity.

2.2.4 LEARN

Learning refers to the ability to acquire knowledge or skills in order to adapt and improve behavior and cognitive understanding (Fiol and Lyles, 1985; Simon, 1981). Learning involves experimentation, model refinement, and integration into products and services, productive processes, and organizational design (Gavetti and Levinthal, 2000; Lawrence and Lorsch, 1967; March, 1991; Senge, 1990). Organizational learning has been a central topic in management research and practice for the past decades (Argote, 2013; Levitt and March, 1988; March, 1991), and machine learning is revolutionizing predictive analytics; it is also core to the development of a vast array of artificially intelligent applications (e.g. Wilson, Sachdev and Alter, 2016). Machine intelligence holds distinctive advantages over

human learning in drawing lessons from large amounts of data as our human information-processing ability is very limited (Simon, 1947, 1973) and prone to serious biases (Kahneman, Slovic and Tversky, 1982). By codifying tacit knowledge, machine learning pushes the boundaries for codified learning, enabling more accurate and scalable learning processes and outcomes. To improve organizations' learning and problem-solving capabilities, humans must identify meaningful problems and shape strategies for acquiring data.

Tesla has used its SCALE capabilities to provide new services to its customers—sometimes as solutions to problems, other times because of suggestions. When the cars were sold in Norway—a new environment—customers complained that cars stopped charging overnight. The cars' computer logs showed this was due to power supply—Norwegian power companies produce electricity with larger variance than the Teslas were calibrated for. A software update slightly widening the security envelope was sent out, and the problem was solved. Similarly, customers complained that it was hard to change the rubber on the windshield wiper—so the company added a "wiper service mode" button to the touch screen with which the wipers sweep up into a vertical position and stay there, providing access.

2.2.5 EXPLAIN

Explaining refers to the ability to show how something works, to explicate causal relationships, articulate purpose, and set direction. It is a leadership imperative but also important in peer-to-peer relationships as well as in interactions with outside parties such as customers, partners, suppliers, and regulators. Explanation is a vital capability in generating organizational purpose, meaning, and identity. The search for explanations drives the identification and formulation of questions and problems for humans and machines to solve.

While machine capabilities on the other four SCALE factors are quite formidable, machines' ability to explain themselves is still very limited, leaving humans primarily responsible for interpretation. This raises a dilemma—the more sophisticated a machine-learning model is, the harder it may be to explain. Explicability is critical for managers' willingness to trust the advice from intelligent systems (Kolbjørnsrud, Amico and Thomas, 2017), though this may change as the tools become more familiar. Furthermore, regulatory initiatives such as GDPR will require organizations to state, in language understandable to customers, how their automated decisions are reached, and to ensure that automated machine learning

does not inadvertently derive unlawfully discriminatory features such as gender or race from the apparent noise of customer backgrounds and behavior.

The technology itself will help, too. Data science technology is increasingly automated for non-data scientists[1] and offers graphical interfaces for a more intuitive analytical process (Schwab, 2018). The latest tools provide better explanation for each individual decision and can be configured to identify the most potent variables. But fully harnessing the power of machine learning may require a reliance on results that are impossible to explain to humans (Weinberger, 2017, 2018), setting us up for a tradeoff between understandable and optimized solutions. If we cannot understand how a model works, we may have to settle for understanding how it behaves.

2.3 HUMAN DECISION MAKING WITH MACHINE LEARNING

> "The Answer to the Great Question... Of Life, the Universe and Everything... Is... Forty-two,' said Deep Thought, with infinite majesty and calm."
>
> — *Douglas Adams (1979), The Hitchhiker's Guide to the Galaxy*

Organizational mastery of the SCALE framework manifests itself in an integrated organizational decision-making process that leverages distinct human and machine capabilities: humans must specify questions comprehensible to machines; machines can then search for solutions to these questions and assess their validity; humans assess whether the machine-generated solutions are viable and valuable; and humans determine deployment procedures.

After expending enormous computational resources, Douglas Adams's Deep Thought answered the humans' question with a laughable answer … because the question was too underspecified. Data-driven "organizational intelligence" obviously requires technological mastery. Perhaps less obviously, data-driven organizations must also engage in cultural change to communicate with machines. Human decision makers must communicate with machines more precisely than they may be accustomed to communicating with other humans. Machines cannot understand the ambiguous or opaque questions that human colleagues sometimes tolerate and muddle through. Human decision makers must learn to ask questions on machines' terms, at a potentially unfamiliar level of precision, and must be able

1. One such tool is DataRobot, which automates much of data scientists' manual work. In most situations, DataRobot's own specialists cannot come up with a model that is better than the one created in autopilot mode.

to assess and manage machines' granular output. Organizations have to master new communication routines to reconcile the "big picture" questions that executives want to answer with the low-level questions machines can answer.

Leaders of successful digital organizations will need to build strategy from questions machines can answer, and to reconcile machine-generated models with the nuance of human preferences. For example, a machine is about as bad at finding "good" customers as it is at finding the meaning of life; but a machine could easily find customers that shop three times a week. Similarly, a machine-learning model might accurately predict that 99.99% of all airline passengers are non-terrorists, but no one will care unless the model can correctly identify the 0.01% of passengers that are.

Imagine that a policy-maker asks the question "How can we reduce cancer-related deaths?" A machine cannot independently answer that question. However, a machine could predict whether a particular tumor is cancerous, a doctor could use that prediction to guide treatment, and that human/machine interaction aggregated over many tumors could reduce cancer-related deaths.

To illustrate the interface between human decision making and machine learning, consider this process. Perhaps a model uses scan measurement to predict whether a tumor is malignant (cancerous) or benign. The question or target "malignant or benign" is a binary problem, easily understood by a machine. With many observations of previous tumors, the machine can use observed cancerous tumors and each tumor's associated scan measurements to train a machine-learning model, resulting in a *confusion matrix*[2]:

TABLE 2.1

Confusion matrix		Predicted outcome	
		Malignant	Benign
Actual outcome	Malignant	31	6
	Benign	1	62

This model is 93% accurate. Is this good? Answering that question depends on what kind of errors we humans care about—and we probably care more about identifying "malignant" than identifying "benign." In economic terms, humans assign different costs to different outcomes. The process of cost assignment

2. The model is generated using SciKit-Learn's decision tree classifier run on the Wisconsin cancer dataset, with the confusion matrix normalized to 100 observations.

remains a uniquely human task: fed precise questions, machines can search for good predictive models, but humans must assess model value.

In this hypothetical example, suppose that correctly identifying a benign tumor is costless. Correctly identifying a malignant tumor triggers a biopsy that costs 1 plus treatment that costs 15. Mistakenly flagging a benign tumor as malignant unnecessarily triggers a biopsy that costs 1. Mistakenly flagging a malignant tumor as benign delays treatment, ultimately triggering a cost of 100. We can show this in a *cost matrix*:

TABLE 2.2

Cost matrix		Predicted outcome	
		Malignant	Benign
Actual outcome	Malignant	16	100
	Benign	1	0

Multiplying the cost matrix by the relative frequencies of each cell in the confusion matrix and summing them up gives an expected cost of each new patient: about 10.97. Should we use the model to decide whether or not to order a biopsy? Probably not—the safe alternative would be to biopsy everyone—at an expected cost of 6.55. However, with a working model and a specified cost function, a human could tweak the model to increase value. Underlying the confusion matrix's binary outcomes are probabilities for each observation: for some of the tumors predicted to be benign, the model was quite sure it was right (say 98% certain); for other tumors predicted to be benign, the model was less confident (say 56% certain). The confusion matrix corresponds to an accuracy-maximizing threshold value separating predicted "benign" cases from predicted "malignant" cases. A human decision maker could reassess this threshold—for instance, only allow the model to predict "benign" if it was more than 90% certain for that instance. This will reduce the model's accuracy but might increase its value by avoiding biopsies for at least some cases.

As automated machine learning becomes more common, human intervention in the modeling process itself will be less necessary, which will facilitate greater machine learning deployment. With increasing deployment of automated machine-learning methods, human intervention in data-driven decision making will increasingly be in the target specification and model evaluation stages. Rather than asking machines to build strategy, humans will need to ask machines ques-

tions that facilitate building strategy from the ground up. Incorporating machines into our decision-making processes by asking them to predict some outcomes will require changes to organizational culture and communication practices.

2.3.1 MODES OF ANALYTICS USE

Organizations vary in how data-driven and intelligent their SCALE capabilities are. We can think of this as variation in the complexity of questions they ask and sophistication in their modeling techniques. Organizations also vary in their pre-analytics capabilities, i.e. collecting and preparing data for analytics, and they tend to be more advanced as they become more experienced and develop skills. There are two main stages of analytics—descriptive and predictive—that can be used to answer different types of questions.

Descriptive analytics is a data-driven approach for questions such as: *What happened?* The analytical focus is to report past and present facts about a situation to human decision makers. It requires few techniques beyond being able to combine, compute and aggregate data, and descriptive statistics. Typical examples are reports and scorecards that report on benchmarks and KPIs. The use of analytics is *passive* to the decision.

Predictive analytics is an analysis-driven approach for question such as: *Why did it happen? What will happen? What should I do?* The focus in these questions shifts from the past to the future, and the application of analytics becomes gradually more advanced.

First, we have what we call a *reactive* mode of analytics in decision making that focuses on understanding why things have happened in the past. In this mode, analytics is used to search for explanations through relationships, patterns and trends. Typical techniques include classical statistics such as cross tabulation, correlation and regression models, but some organizations also apply more advanced techniques such as data mining.

At the next level, analytics are used more directly in decisions, in an *active* mode of decision making where the analysis is aimed at trying to foresee what might happen. The analytical focus is to develop *predictive* models used to estimate probabilities for individual cases (e.g. scoring) and forecasts on aggregated levels. A typical approach is to build predictive models that combine classical statistics and machine learning. Examples include credit scores, fraud identification, and customer intent.

The most advanced level is to deploy the predictive models to guide decisions and actions. We call this a *proactive* mode of analytics in decision making. This

mode has been referred to as prescriptive analytics, but in our view the main difference between the active and proactive modes concerns organizational capabilities to deploy the results from the active mode rather than the techniques applied. In the proactive mode, the models built in the reactive mode, and the probabilities calculated in the active mode, are used to take (or suggest) actions based on rules, simulation or optimization. Examples of applications include smart buildings and server room management (Evans and Gao, 2016). Current applications rely on dynamic and complex networks of automation, often combining a variety of data flows from sources such as transactions, process control systems and sensors, but as organizations become more sophisticated in exploiting the data, this is a changing landscape.

Organizations with capabilities in traditional business intelligence and analytics have focused on descriptive analytics. They need to acquire competencies and skills for predictive analytics. Some have to start with more advanced classical statistics in order to be able to shift from passive to reactive mode. The more advanced levels require the organization to move into data science and learn how to combine techniques such as data management and machine learning with the ability to be able to formulate business problems that the machine can answer. Many organizations will find the shift from descriptive analytics to advanced modes of predictive analytics to be discontinuous—even dramatic.

2.3.2 THE DISCIPLINE OF STRATEGIC EXPERIMENTATION

Increasingly, organizations have to deal with fast-paced and unpredictable change due to new technologies, business models, disruptive innovation, global competition and more. In rapidly and unpredictably changing environments, strategic planning is a risky business. If conditions change even the best plan, once implemented, is likely to be wrong. Such environments require more adaptive approaches to strategy. Organizations armed with strong SCALE intelligence capabilities have the aptitude for strategic, data-driven experimentation (Thomke and Manzi, 2014), as the Norwegian companies RiksTV and Finn.no exemplify.

RiksTV, the Norwegian provider of Digital Terrestrial TV (DTT), faced challenges from bigger competing TV distributors as well as from disruptive over-the-top providers such as Netflix, an inferior distribution platform with limited bandwidth and no on-demand capabilities, and a frequency license expiring in 2021. Management realized that it faces a major digital transformation under great uncertainty and that static, long-term plans would be insufficient. Recognizing the impossibility of specifying a winning strategy ex ante, they frame strategy as a

portfolio of hypotheses that need to be generated, tested, and organized for maximum agility. Inspired by Lean Startup and agile methods, RiksTV continuously develops, executes, and evaluates fast, low-cost experiments in new products and services, as well as technological and operational projects. The experiments are guided by a clear strategic direction specified in four broad goals revised at regular intervals. The experimental turn in the company's strategy has enabled more rapid product development adapted to user needs and preferences, while keeping the organization nimble and capital investment levels moderate.

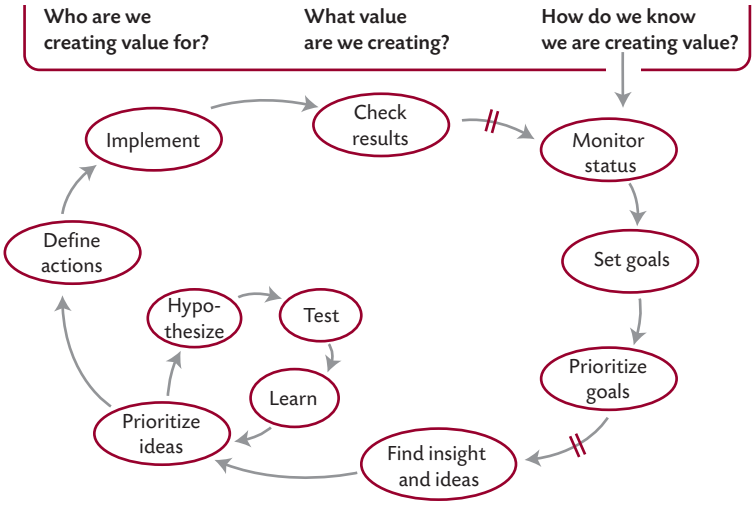

FIGURE 2.1 How they work at Finn.no. Source: Lome, 2016

Finn.no, the online marketplace owned by Schibsted ASA, have explicated a managerial decision process that exemplifies how a data-driven organization changes managerial responsibilities (Lome, 2016). In this process, top management identifies the customer set, the value the organization is creating, and how this value should be measured. After management sets performance goals, they give a development team responsibility for finding a solution. The development team then generates possible solutions and tests them on the digital user interface (a process known as A/B testing.) Solutions that work are implemented and the results of the implementation are checked.

The important change is that solution selection is determined by performance in experiments at the customer interface—not by senior management's judgment (i.e., HIPPO, or "highest paid person's opinion"). Top management may change the goals of the organization or major direction of its activities—which it has

done, notably in switching development from a focus on web interfaces to a "mobile first" strategy—but the detailed implementation within the broader frame of goal-setting is up to the development teams.

The data-driven, intelligent SCALE capabilities enable Finn.no and RiksTV to take experimental yet disciplined approaches to strategizing under uncertainty. Experimentation at SCALE breaks down the traditional strategy formulation and execution divide as formulation and execution are performed in multiple, parallel, and iterative micro-cycles rather than the conventional linear, sequential approach.

2.4 IMPLICATIONS

The emergence of data-driven organizations that intelligently orchestrate collectives of intelligent people and intelligent machines has profound implications for managerial practice, research, and education.

2.4.1 MANAGERS

> "Rather than giving orders as from one person to another, both should take their orders from the situation; justification of order this way is most effective in all situations except when crisis is imminent, in such a case direct order-giving is not only accepted, but expected."
>
> *– Mary Parker Follett, 1941*

Machine learning will, for many organizations, trigger a rethink of management's role. Some decisions will be automated, while in other decisions managerial judgment will be augmented by intelligent technology (Kolbjørnsrud, Amico, and Thomas, 2016; Daugherty and Wilson, 2018), allowing for decisions on how to solve customers' problems to be executed by decentralized resource mobilization, and for rapidly reconfiguring value networks based on customers' interactions with each other.

Reorienting managerial decision making from opinions to data requires discipline from the top. Former Harvard Business School professor Gary Loveman, who built the world's largest casino corporation by changing management decisions from intuition to analysis by carefully analyzing what customers *really* did, famously said that there were three ways to get fired from his company: theft, sexual harassment, and running an experiment without a control group (Schrage, 2011).

Paradoxically, the increased reliance on data and algorithms in decision making and the large-scale automation of routine and information-intensive tasks will increase the need for interpersonal leadership skills among managers, rather than the contrary. The remaining human tasks will be oriented towards creative, complex problem solving requiring managers to harness the collective creativity, intelligence, and judgment of their human coworkers (Chamorro-Premuzic, Wade and Jordan, 2018; Kolbjørnsrud, Amico and Thomas, 2016). The data-driven organization requires bilingual managers that speak both 'machine' and 'human'—that is, know how to communicate and work effectively with both intelligent machines and intelligent humans.

2.4.2 ORGANIZATIONS

According to David (1990), the second industrial revolution's dramatic productivity jump did not emerge from technological change—i.e., transmitting power using electrical cables rather than belts and pulleys—but from recognizing that machines no longer had to line up according to where the belts and pulleys had been. A similar recognition will have to take place in order to fully benefit from the digital revolution—we have to stop designing processes based on the working speed and communications capabilities of individual decision makers, and start thinking of organizations as information processing systems with humans and machines both doing what they do best.

Modern IT systems are organized as individual units that communicate using a common protocol—a computer science concept called "object orientation," first developed in Norway by Dahl and Nygaard (1966) and articulated at Xerox PARC in the 1980s (e.g., Goldberg and Robson, 1983). An important organizational principle is that, as far as possible, one part of the program (one "object") should be used by everyone. Not only does this reduce complexity (if you have an error, you know where to look) but it also ensures that if you come up with a better way of doing something (for instance, a faster way of sorting a list of items) the increased speed will be felt everywhere in the system, since everyone is using the same mechanism.

That is where the difficulty will arise: Models force organizational changes by resolving interdependencies through SCALE. For example, Uber and Lyft invaded the taxi industry by using powerful information technologies to centralize (and take over) some activities (i.e. ordering, communication, pricing, payment, location, navigation and driver/passenger evaluation), while decentralizing other aspects (service design, performance) to the individual driver. The traditional

companies (taxi services) were largely left with financing the vehicles, not as much outcompeted as made obsolete because the new service made most of their managerial decisions based on models and monitoring—and self-organized the rest.

Applying the object-oriented principles to organization design allows the data-driven organization to become actor oriented and self-organized. Its organization design is embedded in the rules and protocols for interaction rather than in a fixed structure, as in hierarchical designs (Fjeldstad et al., 2012; Kolbjørnsrud, 2017). Rules-in-use are the formal and informal rules regulating behavior, rights, obligations in a community—what participants can, cannot, and may do (Crawford and Ostrom, 1995). Protocols are used to guide the interactions of self-organizing actors (Fjeldstad et al., 2012). Combined with extensive information transparency and shared resource commons, the protocols enable a shared situational awareness that allows self-organizing actors—human or machine—to make informed decisions and actions towards fulfilling the goals of the organization—exhibiting distributed intelligence at SCALE.

2.4.3 SCIENCE AND EDUCATION

> But faced with massive data, this approach to science—hypothesize, model, test—is becoming obsolete.
>
> *Anderson (2008)*

Machine learning and intelligent enterprises challenge not just managers, but also scientists. One challenge is that academia is no longer in the lead in the development and application of new methods. Research budgets of large companies dwarf those of universities—companies like Google, Facebook and Baidu develop the new methods for analyzing and acting on data, but also share their findings via open source agreements (see Snow et al., 2017, for an academic treatment, and Dowling, 2017, for a practical example).

In machine learning, there is less use for theory (Anderson, 2008), challenging—at least on the surface—the idea of the scientific method as hypothesis falsification (Popper, 1959). Every Ph.D. student has the proper order of things drilled into them: First you formulate hypotheses from theory, then you collect the data, then you test the hypotheses against the data, and if your coefficients merit three asterisks, you can publish. Doing it in any other order is frowned upon and considered "fishing."

Once overall goal setting is done, however, machine learning is nothing but fishing. Much like managers, researchers doing machine learning will have to

cede model design to the data. This is partially due to increases in computational power: Computers now can test any possible hypothesis (given that you have data) with any known technique (algorithm) and with several competing assumptions to find the "optimal" fit. This approach—raw power over ingenuity and theoretical insight—is something no human can do in their lifetime. Furthermore, at the point of deployment, the computer can continuously test the extent to which the model remains optimal, and continuously adjust it. However, the sheer quantity of data also challenges the very consideration of what is a good model. In social science, most evaluations of models are based on a starved data set: With 300 survey answers, a 95% confidence interval seems a reasonable criterium. If you have millions of observations, everything is significant.

We can view machine learning as large-scale, machine-based induction, developing insights from patterns identified in the data. Machine learning is thus consistent with qualitative researchers' methodological norms. But because qualitative researchers typically lack the skills required to apply machine learning in their work, the scientific community faces a norms/skills paradox: quantitative researchers may have the skills to use machine learning, but machine learning violates their norms about how science should be "done"; qualitative researchers may be open to the approach, but they do not have the required skills. Perhaps this paradox explains why academia has been slower than industry to apply machine learning in empirical research. Going forward, it is imperative that educators teach students the new skills and ways of thinking—to train both next generation scientists and machine learning practitioners.

2.5 CONCLUSION

"There will be no resolution."

(Høeg, 1993)

Ever since Turing (1949) considered whether machines could have intelligence and introduced what came to be known as the Turing test, decision makers have been fascinated with the idea that machines somehow will make man more intelligent. As computers get faster and faster, we have gradually understood that the term artificial intelligence (AI) is somewhat meaningless—as Marvin Minsky is alleged to have said: "Artificial intelligence is anything we haven't done yet."

AI will not make us smarter, and certainly will not replace managers or management decision making anytime soon. It may, however, make managers a bit less prone to biases or, at least, willing to question them; it may make organiza-

tions less wasteful and perhaps more agile; and it may guide the social sciences toward increased relevance. Releasing its potential will require new ways of organizing management, organizations, and science, allowing faster and more precise interactions between all three and their environment. That is to be welcomed—though we still do not know what this will look like, we suggest the technology will allow organizational intelligence to SCALE.

REFERENCES

Adams, D. (1979) *The Hitchhiker's Guide to the Galaxy*. London, UK: Pan Books.

Andersen, E., and Sannes, R. (2017). Hva er digitalisering? *Magma, 2017*(6), 18–24.

Anderson, Chris (2008). The end of theory. *Wired* 2008.06. Retrieved from https://www.wired.com/2008/06/pb-theory/

Andersen, M. N. (2018) *Juice-komet: Vi runder 1 milliard i år*. Børsen. Retreived from http://borsen.dk/nyheder/avisen/artikel/11/189202/artikel.html

Argote, L. (2013). *Organizational learning: Creating, retaining and transferring knowledge* (2nd ed.). New York: Springer.

Brynjolfsson, E., and McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W.W. Norton.

Chamorro-Premuzic, T., Wade, M., and Jordan, J. (2018). As AI makes more decisions, the nature of leadership will change. *Harvard Business Review*. Retrieved from https://hbr.org/2018/01/as-ai-makes-more-decisions-the-nature-of-leadership-will-change

Crawford, S. E., and Ostrom, E. (1995). A grammar of institutions. *American Political Science Review, 89*(3), 582–600.

Dahl, O. J., and Nygaard, K. (1966). *SIMULA: An Algol Based Simulation Language*. Oslo, Norway: Norwegian Computing Center.

Daugherty, P. and Wilson, H. J. (2018) *Human + Machine: Reimagining Work in the Age of AI.* Boston, MA, Harvard Business Review Press

Davenport, T. H., and Kirby, J. (2016). *Only humans need apply: winners and losers in the age of smart machines.* Harper Business.

Davenport, T. H. and H. H. Kim (2013). *Keeping Up with the Quants: Your Guide to Understanding and Using Analytics*. Boston, MA, Harvard Business Review Press.

David, P. A. (1990). The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox. *American Economic Review*, 80(2), 355–361.

Denning, P. J. and T. G. Lewis (2016). "Exponential laws of computing growth." *Communications of the ACM* 60(1): 54–65.

Dowling, J. (2017) Distributed TensorFlow. *O'Reilly Radar*, December 19. Retrieved from https://www.oreilly.com/ideas/distributed-tensorflow

Evans, R., and Gao, J. (2016, 06.20.2016). DeepMind AI Reduces Google Data Centre Cooling Bill by 40%. Retrieved from https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/

Fiol, C. M., and Lyles, M. A. (1985). Organizational learning. *Academy of Management Review, 10*(4), 803–813.

Fjeldstad, Ø. D., and Snow, C. C. (2017). Business models and organization design. *Long Range Planning*. doi: https://doi.org/10.1016/j.lrp.2017.07.008

Fjeldstad, Ø. D., Snow, C. C., Miles, R. E., and Lettl, C. (2012). The architecture of collaboration. *Strategic Management Journal, 33*(6), 734–750. doi:10.1002/smj.1968

Follett, M. P. (1941). *Dynamics of Administration*. New York: Harper and Row.

Galbraith, J. R. (1973). *Designing Complex Organizations*. Reading, MA.: Addison-Wesley.

Galbraith, J. R. (2010). The multi-dimensional and reconfigurable organization. *Organizational Dynamics, 39*. doi:10.1016/j.orgdyn.2010.02.001

Gavetti, G., and Levinthal, D. (2000). Looking forward and looking backward: Cognitive and experiential search. *Administrative Science Quarterly, 45*(1), 113–137. doi: 10.2307/2666981

Goldberg, A., and Robson, D. (1983). *Smalltalk 80—the Language and Its Implementation*. Reading, MA: Addison-Wesley.

Google. (2017). The story of AlphaGo so far. Retrieved from https://deepmind.com/research/alphago/

Høeg, P. (1993) *Miss Smilla's Feeling for Snow*, Harvill Press, UK.

Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*: Cambridge University Press.

Kolbjørnsrud, V. (2017). Agency problems and governance mechanisms in collaborative communities. *Strategic Organization, 15*(2), 141–173. doi:10.1177/1476127016653727

Kolbjørnsrud, V., Amico, R., and Thomas, R. J. (2016). How Artificial Intelligence Will Redefine Management. *Harvard Business Review*.

Kolbjørnsrud, V., Amico, R., and Thomas, R. J. (2017). Partnering with AI: How organizations can win over skeptical managers. *Strategy and Leadership, 45*(1), 37–43.

Lawrence, P. R., and Lorsch, J. W. (1967). *Organization and environment: Managing differentiation and integration*. Boston: Division of Research, Graduate School of Business Administration, Harvard University.

Levitt, B., and March, J. G. (1988). Organizational learning. *Annual Review of Sociology, 14*, 319–340. doi: 10.1146/annurev.so.14.080188.001535

Lome, K. B. (2016) How we work: Working to ensure user value, innovation and performance. *Schibsted Bytes*. Retrieved from http://bytes.schibsted.com/working-ensure-user-value-innovation-performance/

Lyche, K. (2017) Roboten "James" hanket inn 30 kunder første året. Dagens Næringsliv. Retrieved from https://www.dn.no/grunder/2017/11/05/1801/Finans/roboten-james-hanket-inn-30-kunder-forste-aret

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71–87.

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics, 38*(8), 4.

Muller, J. (2013) Tesla's Gotcha Blog Catches New York Times Reporter Driving In Circles, *Forbes*, February 14, retrieved from https://www.forbes.com/sites/joannmuller/2013/02/14/teslas-gotcha-blog-catches-new-york-times-reporter-driving-in-circles/#756e02b4fe73

Popper, K. (1959). *The Logic of Scientific Discovery*. New York, Basic Books.

Sasson, A., and Johnson, J. C. (2016). The 3D printing order: variability, supercenters and supply chain reconfigurations. *International Journal of Physical Distribution and Logistics Management, 46*(1), 82–94.

Schrage, M. (2011) Q&A: The Experimenter. *MIT Technology Review*, March/April, retrieved from https://www.technologyreview.com/s/422784/qa-the-experimenter/

Schwab, K. (2018) This is the world's first graphical AI interface. *Co.Design*, retrieved from https://www.fastcodesign.com/90157777/this-is-the-worlds-first-graphical-ai-interface

Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization* (1st ed.). New York: Currency Doubleday.

Simon, H. A. (1947). *Administrative behavior: A study of decision-making processes in administrative organizations*. New York: Macmillan.

Simon, H. A. (1973). Applying information technology to organization design. *Public administration review, 33*(3), 268–278.

Simon, H. A. (1981). *The sciences of the artificial*. Cambridge, MA: MIT Press.

Simon, H. A., and Newell, A. (1958). Heuristic problem solving: The next advance in operations research. *Operations research, 6*(1), 1–10.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature, 550*, 354. doi:10.1038/nature24270. Retrieved from: https://www.nature.com/articles/nature24270

Snow, C. C., Fjeldstad, Ø. D., and Langer, A. M. (2017). Designing the digital organization. *Journal of Organization Design, 6*(1), 7.

Sterling, B. (2017). Deep Mind Alpha Zero's "Immortal Zugzwang Game" against Stockfish. *Wired,* retrieved from https://www.wired.com/beyond-the-beyond/2017/12/deep-mind-alpha-zeros-immortal-zugzwang-game-stockfish/

Teece, D. J., Pisano, G., and Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7), 509–533. doi:10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z

Thomke, S., and Manzi, J. (2014). The discipline of business experimentation. *Harvard Business Review, 92*(12), 70–79.

Tushman, M. L., and Nadler, D. A. (1978). Information processing as an integrating concept in organizational design. *Academy of Management Review, 3*(3), 613–624.

Turing, A. M. (1949). Computer Machinery and Intelligence. *Mind, 59*(236), 433–460.

Waters, R. (2018, 01.28.2018). Microsoft and Google prepare for big leaps in quantum computing. *Financial Times*. Retrieved from https://www.ft.com/content/4b40be6c-0181-11e8-9650-9c0ad2d7c5b5

Weinberger, D. (2017). Our machines now have knowledge we'll never understand. *Wired*. Retrieved from: https://www.wired.com/story/our-machines-now-have-knowledge-well-never-understand/

Weinberger, D. (2018). Don't make AI artificially stupid in the name of transparency. *Wired*. Retrieved from https://www.wired.com/story/dont-make-ai-artificially-stupid-in-the-name-of-transparency/

Wilson, H. J., Daugherty, P. R., and Morini-Bianzino, N. (2017). The Jobs That Artificial Intelligence Will Create. *MIT Sloan Management Review, 58*(4).

Wilson, H. J., Sachdev, S., and Alter, A. (2016). How companies are using machine learning to get faster and more efficient. *Harvard Business Review*.

Winston, P. H. (1984). *Artificial Intelligence*. Reading, MA: Addison-Wesley.

Wood, G. (2009). #IPASocial Principle 9: Change will never be this slow again. *GeekMedia*. Retrieved from http://graewood.blogspot.no/2009/09/ipasocial-principle-9-change-will-never.html

DOI: 10.18261/9788215031583-2018-04

Chapter 3
Making the Digital Transformation Work

SUT I WONG AND CHRISTIAN FIESELER

ABSTRACT In this chapter, we highlight the on-going research of BI Norwegian Business School's Nordic Center for Internet and Society to better understand the function, status, and meaning creation of work in the digitized economy, and the impact of digital technologies in organizations. Specifically, the chapter aims to set out an agenda for mastering the labor challenges of the digital transformation based on our studies conducted over recent years. We highlight the challenges of adapting our current notions of managerial feedback to platform organizations, and present insights on how to support the creative potential of online crowdsourcing. Further, we showcase the pitfalls of the emerging practice of virtual leadership and propose measures with which good leaders may greatly increase the effectiveness of online communities. Lastly, we conclude by outlining what might constitute attractive organizations for the future workforce and labor designs that could render the digital economy more inclusive, effective, and human-centered.

KEY WORDS: Future Workplaces | New Working Modes | Labor Designs | Crowdworking | Virtual Teams | Digital Leadership

3.1 INTRODUCTION

A bestseller published by Erik Brynjolfsson and Andrew McAfee (2014) titled *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* is one instance in a long list of recent publications exploring the effect of new information and communication technologies on labor markets. New technologies such as artificial intelligence employing self-learning algorithms and large bodies of data will alter future collaboration between humans and machines. Observers posit that such digital technologies will enable the automation of industrial as well as cognitive tasks previously reserved for human ingenu-

ity, triggering an era of accelerated innovation and significant disruption comparable to the fourth industrial revolution (Frey and Osborne, 2013).

While new information and communication technologies may facilitate the substitution of human labor, they also allow for the emergence of new forms of work. A closer look at the emerging on-demand service economy, for instance, reveals a growing workforce characterized by commodification, low costs, minimal institutionalization, and increasing anonymity. Digital platforms and ecosystems such as Upwork or Topcoder, and even sharing platforms like Airbnb and Uber, bring disruptive change to existing industries by enlisting the work of thousands of dispersed individual workers (Kneese and Rosenblat, 2014). The traditional model of labor is hence under attack from two directions simultaneously, through replacement by new technologies, and through commodification facilitated through new, technology-enabled forms of organizing. We will call the new work model that will emerge in its place "digital work."

As technological innovations disrupt traditional forms of employment, and new forms of labor emerge, we have yet to develop a thorough understanding of how the nature and meaning of digital work will evolve in the future. Some researchers, taking an optimistic view, have pointed to the potential benefits of online micro-entrepreneurship, such as flexibility, enjoyment, and the economic empowerment of previously constrained individuals (Fish and Srinivasan, 2012; Gansky, 2010; Horten, 2011; Kelliher and Anderson, 2009; Kneese and Rosenblat, 2014; Ruggieri, Mosconi, Poponi, and Silvestri, 2015).

Conversely, digital work is, at least to date, often considered to be remote, modular, and conducted on a project-by-project basis, limiting the creation of permanent ties to employers, organizations, or co-workers (Andrejevic, 2009; Ashford et al., 2007; Connelly and Gallagher, 2004; de Peuter, 2011; Fuchs and Sevignani, 2013; Gill and Pratt, 2008; Gregg, 2011; Hollister, 2011; Horowitz and Rosati, 2014; McKercher and Mosco, 2008; Rainie and Wellman, 2012; Smith, 2016). Digital workers laboring on decontextualized projects in comparatively social isolation are still often merely viewed as outsourced "human computers." Nevertheless, the nature and meaning of one's work accounts for a critical part of the identity of many people (Blustein, 2011). In other words, they retain the natural human desire to feel needed, to feel valued, and to have their work appreciated by the community and larger society (Jung, 2015). In short, they want to feel that they "matter."

3.2 THE ORIGINS AND CHALLENGES OF THE DIGITAL TRANSFORMATION FROM AN INDUSTRIAL RELATIONS PERSPECTIVE

In the past decade, much attention has been paid to the ongoing development of the fourth industrial revolution, which is set to change work, work practices, and workplaces (Colbert, Yee, and George, 2016). In regard to the transformation of work, this revolution is sought either to fundamentally alter or outright replace existing work, such as many clerical professions, or to be conducive to large-scale projects that were formerly the sole purview of more traditional forms of organizations, which are to be broken down into small work packages that can be distributed among a digitized workforce (Lehdonvirta and Ernkvist, 2011; Kittur et al., 2013). Currently, this involves more menial tasks such as usability testing, image tagging, audio transcription and/or evaluation, and text fragment categorization, but also encompasses the gig economy that provides, for instance, transportation and hospitality services. However, with improving technology and organizational design, it will also increasingly include creative and innovative tasks. Progressively greater numbers of individuals are either making a living or earning additional income through freelance contracting on the Internet. Examples of this include the completion of human-intelligence tasks on Amazon Mechanical Turk (AMT) and Clickworker, and the offering of software development or design skills via crowdsourcing platforms such as Upwork or 99designs.

The overall size of the digital gig economy was estimated at US$2 billion in 2013 and grew to US$4.8 billion in 2016 (Kuek et al., 2015). The emergence of the sector has been driven by the platforms' combination of competitive logic and technical innovation, which they have used not only to win market share from existing IT outsourcing, but also to generate new forms of outsourcing (Huws, 2017). Growth has been further driven on the supply side by the competitive logic of clients seeking and obtaining three key benefits: lower costs (financial and time), greater flexibility, and access to a wider skills pool (i.e. higher quality workers) (Bergvall-Kareborn and Howcroft, 2014).

Many in traditional employment may also witness changes in the nature and meaning of their work as a result of the recent development of digital technologies. For instance, with digital technologies foreseeably automating many incumbent forms of employment, some may face a transition of their formal work roles (e.g., as they are replaced by or have to manage robots and artificial intelligence for manufacturing) (Wolf, 2016). Indeed, the number of jobs that largely rely on routine tasks is predicted to decrease (Hilton, 2008). New employment is predominately arising at the fringes of the traditional labor market, such as in the afore-

mentioned gig and freelance economy. As computers are still not good at abstract tasks, which often require higher skills, and manual tasks with lower skills requirement, skills and competencies required in the future are said to be increasingly polarized (Hilton, 2008). Moreover, the meaning of work is also changing for many. For example, nursing has traditionally been seen as a hand-holding profession. However, since nurses' work is being increasingly digitized, their work identity has been suggested to be moving away from the humanized aspect toward technical skills (Kolbæk, 2015).

In recent years, a rich body of literature has emerged that tries to shed light on the nature of this digital transformation. The literature itself is divided into several scientific disciplines, discourses, and theoretical approaches. Exemplary disciplines involved in the study of such new forms of work and work practices include sociology/anthropology (Fish and Srinivasan, 2012; Pinch and Bijker, 2012), communication and media studies (Irani, 2015; Paul M Leonardi, 2015; Martin, Parry, and Flowers, 2015; Sarker, Ajuja, Sarker, and Kirkeby, 2011), psychology (Brown, Venkatesh, Kuruzovich, and Massey, 2008; Hoch and Kozlowski, 2014), organization studies (Bauer and Gegenhuber, 2015; Boons, Stam, and Barkema, 2015; Kirkman, Rosen, Tesluk, and Gibson, 2004; Spreitzer, Cameron, and Garrett, 2017), and information systems and computer science, such as computer-supported cooperative work and human–computer interaction (Kittur et al., 2013; Venkatesh and Goyal, 2010).

In the following discussion of our own research, we postulate that digital technologies have transformed labor into three distinct forms, leading to three distinct types of labor, namely aspirational, platform, and corporate labor, as illustrated in Figure 3.1. At the current point in time, corporate labor is still the form of labor that, financially, allows a career to be sustained the best; aspirational labor (as will be explained in more detail below) currently pertains more to career entry activities, particularly in the creative industries; whereas platform labor takes a middle ground encompassing increasingly more fluid freelancing agreements. We expect all three forms of labor to become increasingly more collaborative, as we have already witnessed in the preceding decades— depicted in Figure 3.1 as one of the fundamental dimensions of future work. By Collaborative Creation we in essence refer to work that is to be increasingly split up into more distinct parts to be worked on either through human or artificial intelligence. As our final dimension, depicted in the figure below as Organizational Openness, we also expect a more open organizational design, in which the boundaries between organizations become increasingly blurred.

Digitization empowers all three types of labor, which are enabled through technologies. For instance, traditional corporate labor can be performed in established settings but in novel ways, with less geographical and time constraints, in order to make a workplace more inclusive. For individuals active in freelancing on platform organizations, digital technologies enable a new type of boundaryless career, where workers and employers are matched on a case-by-case basis through a digital intermediary—the platform. Finally, there is the mostly unremunerated creative, aspirational labor that, nonetheless, may be not only critical to the value creation of companies, but also a means to pursue one's hobby and/or passion in a rewarding career.

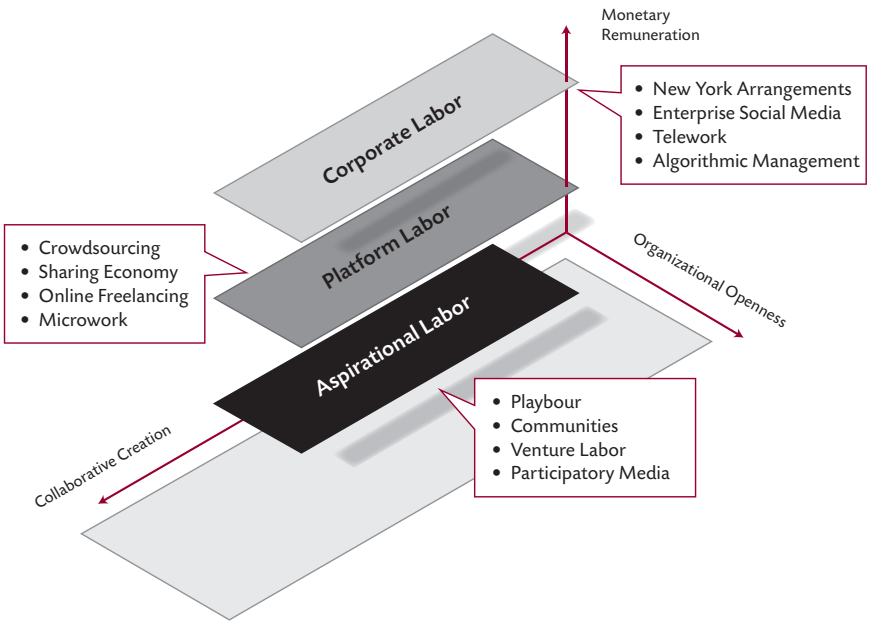

FIGURE 3.1

3.2.1 ASPIRATIONAL LABOR

Although online waged labor is a central construct within the digital economy, digital laborers can also be motivated by numerous non-monetary factors, such as social influence and hedonism, resulting in work being offered on a spectrum of paid to unpaid "free labor" (Boons, Stam, and Barkema, 2015; Manyika et al., 2014; Terranova, 2004). Increasingly, the boundaries of labor and play are blurring, with the neologism "playbor" emerging to describe work practices that,

according to Scholz (2013, p. 3), "don't feel, look, or smell like labor at all." In the face of a flexible and entrepreneurially driven digital economy, more and more people desiring career entry into desirable digital media professions are being driven toward unpaid online activities such as blogging, gaming, and branding. These activities are undertaken for enjoyment, but also for the development of skills and networks. For many, these activities are their creative outlets and primary method of making new friends with similar interests. More cynically, this phenomenon of "playbor" is merely work "suffused with an ideology of play, which effectively masks labor as play, and disguises the process of self-expropriation as self-expression" (Kücklich, 2009). Yet, a side effect of many of these activities may be the generation of income, variable in amount but with the potential to result in considerable sums of money. Not only has this financial gain materialized through immediate returns, such as shared advertising revenue on YouTube videos or selling handmade items on Etsy, but some people are also viewing their online activities as investments for future rewards. The development of networks, skills, and online identities tied to desirable industry sectors has enabled people to increase their employment opportunities in the future. We see this, for instance, in how an aspiring teen blogger writes daily on her food blog in the hope of becoming a professional journalist, or how a young Twitch star posts videos of increasingly higher quality in the hope of leveraging that experience for a position in a traditional media or video game company. This ostensibly "free" labor online is thus being bargained for the hope of future payoff, a phenomenon entitled "hope labor" (Kuehn and Corrigan, 2013) or "aspirational labor" (Duffy, 2016).

3.2.2 PLATFORM LABOR

This type of independent contract functions as a direct relationship involving just two parties: the worker and the client organization (s)he contracts with. From task selection to completion, the work process is controlled by the worker; an independent contractor's only employer is him/herself, and each project has a relatively short time span (Deng and Joshi, 2016). This constitutes a new form of employment in which there is a great deal of flexibility in the employment relationship even compared with traditional freelance work (Spreitzer et al., 2017). All platform workers have some autonomy in terms of controlling their own scheduling and where the work is done. Moreover, the allocation of decision-making authority across workers, clients, and the firm varies substantially across platforms, as does the degree to which workers are compensated according to outputs or inputs.

The affordances of online platforms enable the offering of high skilled and creative labor of the type that tends to be described as both satisfying and pleasurable (Gill and Pratt, 2008). Some scholars have even characterized the digital economy as a fertile environment for democratic free production that enables individuals to express their creativity and transcend alienation (Bruns, 2008; Postigo, 2016; Jenkins, 2006; Tapscott and Williams, 2006; Howe, 2009; Florida, 2002; Prat and Gill, 2000; Hesmondhalgh and Baker, 2011). However, in the cases of highly skilled "creative labor" and "digital entrepreneurship," the levels of digital literacy may restrict access to the online platforms. The prerequisite for participation can therefore be high for those who are already imbued with offline economic, social, and cultural capital.

Conversely, with new technologies enabling the differentiation, specification, and outsourcing of labor, low-skilled forms of labor, such as crowdwork, microwork, and digital piecework, have become increasingly common (Ashford et al., 2007; Fish and Srinivasan, 2012; Kittur et al., 2008, 2013; Kneese and Rosenblat, 2014; Lehdonvirta and Ernkvist, 2011; Postigo, 2016; Silberman et al., 2010). Criticism has been leveled at the affordances of this model to increasingly fragment and substitute previously middle-class jobs as technologically induced competitive forces lead to the overall deterioration of wages and working conditions (Scholz, 2013).

Despite the growth of the digital economy across high-skilled, low-skilled, and so-called playbor manifestations, the organizational understanding of what motivates digital workers beyond financial compensation is still largely limited (Chua, Roth, and Lemoine, 2015; Kosonen, Gan, Vanhala, and Blomqvist, 2014). Given this lack of understanding, it is easy for employers to misconstrue the digitized workforce as an amorphous crowd of exchangeable workers instead of a valuable community of individuals, each with differing motivations and experiences (Kittur et al., 2013). Moreover, as workers become increasingly dependent upon established platforms as entry points into the digital economy, the nature and affordances of such platforms dictate the type, frequency, reward system, and context of digital work (Fuchs and Sevignani, 2013; Kingsley, Gray, and Sury, 2015; Scholz, 2013; Rosenblat and Stark, 2015). Without insight into worker experiences, digital organizations may be perpetuating unfair labor conditions that ignore the human element of their workforce. On the other hand, without oversight, digital platforms retain the ability to exploit workers through unfair power dynamics while simultaneously alienating them from their own intellectual products (Arvidsson, 2008; Aytes, 2013, Brabham, 2008; 2012; Fuchs, 2010; Kalekin-Fishman and Langman, 2015; Postigo, 2016; Terranova, 2004; van Dijck and Nieborg, 2009; Zwick, Bonsu, and Darmody, 2008).

3.2.3 CORPORATE LABOR

Many organizations have introduced technologies for changing the organizational processes in traditional work settings, such as using information technology in hiring processes, enterprise social media as communication platforms, new working arrangements such as office design (e.g., flexible seating enabled by cloud technologies) and telework (e.g., flexible work locations and scheduling relying on computer-mediated communication tools) (Colbert, Yee and George, 2016). Work has become more flexible in terms of both scheduling and location (Spreitzer et al., 2017) and potentially more social (Paul M Leonardi, 2015) with the use of new technologies.

Often, technology is adopted with an intention to enable better efficiency and work conditions. For instance, Enterprise Social Media (ESM) is said to provide numerous benefits for organizations by making communication less bureaucratic and more transparent and inclusive (Leftheriotis and Giannakos, 2014). Some argue that ESM, as a potential way of displaying work behaviors, attitudes, and organizational culture, may help organizational members not only acquire explicit knowledge, which refers to articulated, expressed, and recorded knowledge such as organizational visions and role descriptions, but also tacit knowledge, which refers to know-how that is more intuitive (Paul M. Leonardi and Treem, 2012). For telework and other flexible work practices enabled by computer-mediated communication tools such as instant chat, video conferencing, online forums, etc., individuals may see that these arrangements would make collaborating and knowledge sharing easier and less bureaucratic (Martins, Gilson, and Maynard, 2004; Sarker et al., 2011). Although technologies are approached as a means to appropriate organizational strategies, the success of such intentions are dependent on the social environment (Dutrénit, 2004; Pasmore, Francis, Haldeman, and Shani, 1982; Powell, 1987)

In particular, while flexible work arrangement are appreciated by some, others may find themselves less attached to the office and/or the organization, or the work environment may become less personal, less social, and more difficult to share knowledge in (Cheshin, Rafaeli, and Bos, 2011; Hertel, Geister, and Konradt, 2005; Mesmer-Magnus, DeChurch, Jimenez-Rodriguez, Wildman, and Shuffler, 2011). Individuals may thus see technology as constraining rather than enabling. This is particularly so when they see technology as a structural property of organizations that reduces the flexibility with which they would go about their work (Orlikowski, 1992). It is therefore not surprising that research findings on the effect of technology adoption on individual performance (Lewis, Agarwal, and Sambamurthy, 2003) and team performance (Ortiz de Guinea, Webster, and Staples, 2012) have been inconsistent.

3.3 FUTURE RESEARCH AGENDA

Clearly, understanding the conditions under which digital technologies are shaping organizational phenomena is an important current research agenda (Ashford, George, and Blatt, 2007; Colbert et al., 2016; Piezunka and Ander, 2015; Spreitzer et al., 2017). Specifically, there is an urgent need to advance our understanding of individual attitudinal and behavioral responses in digitized workplaces (Boons et al., 2015; Gibson and Gibbs, 2006; Nakatsu, Grossman, and Iacovou, 2014). Both public institutions and private organizations share responsibility in creating a fair digital economy for workers. Accordingly, by focusing on both procedural and interactional fairness, our goal is to provide evidence-based implications for policy makers and stakeholders of what factors or procedures may support positive work arrangements for digital workers. The dependence of workers on online platforms raises important questions as to organizational power dynamics. Corporate and institutional social responsibility will thus be a key component within this research question.

For instance, some of our empirical studies demonstrate that platform labor may see the instant digital feedback they receive as the result of surveillance rather than support. Considering that such instant feedback is indeed a primary social stimulus for platform laborers, as they are working through the mediation of an online platform without face-to-face or other means of interaction (Gamrat, Zimmerman, Dudek, and Peck, 2014), this has important implications not only for their performance, but also for their psychological well-being. We have also observed that platform laborers do look for meaning in their work and hope to see their work matter, not least so they can develop their careers in platform work, despite the lack of vertical career mobility and competence development. We will pay particular attention to the role of worker voice, feedback mechanisms, and the provision of development opportunities for workers. We also want to question the responsiveness of platforms to the changing needs of workers. As greater numbers of companies transition to a partly or wholly digital workforce, we question whether organizations are evolving with worker needs in mind or are even redefining what it means to be an "employee" for internal benefit. With this research question, we hope to generate cutting-edge empirical research that can help different stakeholders to identify exclusionary elements, leading to the creation of mechanisms that will foster greater inclusion and fairness.

Following this research question, we also aim to examine the role of entry into the digital economy and consider how aspiring digital workers acquire the necessary digital literacy skills to engage online. With the rise of online work, traditional jobs are increasingly being replaced by portfolio careers full of diverse

activities. With our contribution, we consider how and why people choose different jobs at different points in their lives, and how the boundaries between those might intersect or conflict. We also consider digital skill divides, which may be perpetuated by offline social and economic factors. Of these, proactive skill development and adjustment in the face of new career demands has gained greater attention, particularly regarding the rapidly evolving nature of the digital economy. Skills developed in educational settings may be misaligned to the needs of the current and future workplace. Accordingly, we will challenge the role of youth engagement and preparation for the future of work, informing stakeholders in education how to provide sensible interventions that spur effective skill development both online and offline.

Furthermore, organizations have increasingly been employing distributed teams with digital solutions as means to improve organizational efficiency and effectiveness (Colbert et al., 2016). For many workers today, their jobs are not confined to a specific location or point in time. On the contrary, teams can easily be arranged across temporal, geographical, and organizational boundaries (Hoch and Kozlowski, 2014). As new ways of working, such distributed teams are assumed to benefit employees and organizations alike, with potential benefits such as increased flexibility, work–life balance, job satisfaction, and performance (Gilson, Maynard, Young, Vartiainen, and Hakonen, 2015; Liao, 2017; Martins et al., 2004). However, despite the alleged benefits, this new team format tends to obtain less desirable individual and organizational outcomes than traditional teams do (Gibson and Gibbs, 2006; Ortiz de Guinea et al., 2012). Distributed teams have been reported to display less extra-role behaviors (Ganesh and Gupta, 2010) and experience more communication issues (Daim et al., 2012; Ortiz de Guinea et al., 2012) and increased task conflict (Ortiz de Guinea et al., 2012). Other disadvantages of distributed teams as opposed to co-located teams include lower levels of satisfaction with work and one's team, as well as less knowledge sharing (Ortiz de Guinea et al., 2012), trust, team cohesion, cooperative behavior, and social control (Hoch and Kozlowski, 2014).

Leading in such digitized workplaces can be challenging, as leaders who are not geographically present typically have a harder time withholding an active, relationship-focused leadership style, as they do not have other means to communicate with their employees aside from computer-mediated communication tools (Dulebohn and Hoch, 2017). Supporting this notion, Huang et al. (2010) found that media richness influences the relationship between transformational leadership and cooperative climate in distributed teams. One of our empirical studies demonstrates that transformational leadership is likely to be less effective in building high-quality

leader–member relationships in teams that rely strongly on electronic communication tools, such as video conferencing, instant messages, phone calls, and emails, to go about their daily work. The weakened effect is particularly likely to occur with members in distributed teams who have high task interdependency among each other. This indicates that the role of leadership is likely to be substituted in distributed teams in which members have a high frequency of interactions among each other. While this is perhaps good news for organizations that aim for a flat structure, for most of the organizations that do have some levels of hierarchical structure, ineffective leadership can be detrimental. Based on this finding, we challenge current management theories based on traditional organizational settings, which may not necessarily apply to the more open and fluid organizational processing enabled by digital technologies, and suggest that more research is needed.

The final research question will act as the culmination of the project, drawing insights from the former research questions. Our overall aim is to integrate various theory and research streams to illuminate the manifold experiences of "mattering" within the digital economy. We want to give the individual workers in the digital economy a voice and listen to what is being said, inviting workers to tell their own stories. How do they feel in the digital workspace? Do workers perceive different experiences of mattering from creative work, unwaged hope labor, and unskilled piece work? What implications does the ability to contextualize one's own work have for individual well-being and productivity? Our goal is to create guidelines, ideas, and hands-on organizational design principles that might provide greater meaning with digital work, improving worker experiences and promoting the acceptance of digital labor formats by potentially uncertain workers. Above all, we want to refocus ongoing debates on individual workers, including aspirational, platform, and corporate laborers, without whom there would be no digital economy.

REFERENCES

Andrejevic, M. (2009). Exploiting YouTube: contradictions of user-generated labor. *The YouTube Reader*, 406–413.

Arvidsson, A. (2008). The ethical economy of customer coproduction. *Journal of Macromarketing*, *28*(4), 326–338.

Ashford, S. J., George, E., and Blatt, R. (2007). Old Assumptions, New Work: The Opportunities and Challenges of Research on Nonstandard Employment. *The Academy of Management Annals*, 65–117.

Aytes, A. (2013). Return of the Crowds: Mechanical Turk and Neoliberal States of Exception. In T. Scholz (ed.), *Digital Labor. The Internet as Playground and Factory,* 79–97. New York: Routledge.

Bauer, R. M., and Gegenhuber, T. (2015). Crowdsourcing: Global search and the twisted roles of consumers and producers. *Organization*, 22(5), 661–681. doi:10.1177/1350508415585030

Bergvall-Kåreborn, B., and Howcroft, D, (2014). Amazon Mechanical Turk and the commodification of labour. *New Technology, Work and Employment,* 29(3), 213–223.

Blustein, D. L. (2011). A relational theory of working. *Journal of Vocational Behavior*, *79*(1), 1–17.

Boons, M., Stam, D., and Barkema, H. G. (2015). Feelings of pride and respect as drivers of ongoing member activity on crowdsourcing platforms. *Journal of Management Studies*, 52(6), 717–741.

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, 14(1), 75–90.

Brabham, D. C. (2012). The myth of amateur crowds: A critical discourse analysis of crowdsourcing coverage. *Information, Communication and Society*, *15*(3), 394–410.

Brown, S. A., Venkatesh, V., Kuruzovich, J., and Massey, A. P. (2008). Expectation confirmation: An examination of three competing models. *Organizational Behavior and Human Decision Processes*, 105(1), 52–66. doi:10.1016/j.obhdp.2006.09.008

Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and beyond: From production to produsage* (Vol. 45). Peter Lang.

Brynjolfsson, E., and McAfee, A. (2014). *The Second Machine Age*. New York: Norton

Cheshin, A., Rafaeli, A., and Bos, N. (2011). Anger and happiness in virtual teams: Emotional influences of text and behavior on others' affect in the absence of non-verbal cues. *Organizational Behavior and Human Decision Processes*, 116(1), 2–16. doi: http://dx.doi.org/10.1016/j.obhdp.2011.06.002

Chua, R. Y., Roth, Y., and Lemoine, J. F. (2015). The impact of culture on creativity: How cultural tightness and cultural distance affect global innovation crowdsourcing work. *Administrative Science Quarterly*, *60*(2), 189–227.

Colbert, A., Yee, N., and George, G. (2016). The digital workforce and the workplace of the future. *Academy of Management Journal*, 59(3), 731–739.

Connelly, G. E. and Gallagher, D. G. (2004). Emerging trends in contingent work research. *Journal of Management*, 30(6), 959–983.

Daim, T. U., Ha, A., Reutiman, S., Hughes, B., Pathak, U., Bynum, W., and Bhatla, A. (2012). Exploring the communication breakdown in global virtual teams. *International Journal of Project Management*, 30(2), 199–212. doi: https://doi.org/10.1016/j.ijproman.2011.06.004

De Peuter, G. (2011). Creative economy and labor precarity: a contested convergence. *Journal of communication inquiry*, *35*(4), 417–425.

Deng, X., Joshi, K. D., and Galliers, R. D. (2016). The duality of empowerment and marginalization in microtask crowdsourcing: Giving voice to the less powerful through value sensitive design. *MIS Quarterly*, *40*(2), 279–302.

Duffy, B. E. (2016). The romance of work: Gender and aspirational labour in the digital media industries. *Journal of Cultural Studies*, 19(4), 441–457.

Dulebohn, J. H., and Hoch, J. E. (2017). Virtual teams in organizations. *Human Resource Management Review*, 27(4), 569–574. doi: https://doi.org/10.1016/j.hrmr.2016.12.004

Dutrénit, G. (2004). Building Technological Capabilities in Latecomer Firms: A Review Essay. *Science, Technology and Society*, 9(2), 209–241. doi: 10.1177/097172180400900202

Fish, A., and Srinivasan, R. (2012). Digital labor is the new killer app. *New Media and Society*, 14(1), 137–152. doi: 10.1177/1461444811412159

Florida, R. (2002). *The rise of the creative class, and how it is transforming work, leisure, community and everyday life*. New York.

Frey, C. B., and Osborne, M. A. (2013). The future of employment: how susceptible are jobs to computerisation? *Oxford University Working Paper*.

Fuchs C and Sevignani S (2013) What Is Digital Labour? What Is Digital Work? What's their Difference? And Why Do These Questions Matter for Understanding Social Media? *tripleC: Communication, Capitalism and Critique*, 11(2): 237–293.

Fuchs, C. (2010), Labor in informational capitalism on the internet. *The Information Society, 26,* 179–196.

Gamrat, C., Zimmerman, H. T., Dudek, J., and Peck, K. (2014). Personalized workplace learning: An exploratory study on digital badging within a teacher professional development program. *British Journal of Educational Technology*, 45(6), 1136–1148. doi: 10.1111/bjet.12200

Ganesh, M. P., and Gupta, M. (2010). Impact of virtualness and task interdependence on extra? role performance in software development teams. *Team Performance Management: An International Journal*, 16(3/4), 169–186. doi: 10.1108/13527591011053250

Gansky, L. (2010). *The mesh: Why the future of business is sharing*. Penguin.

Gibson, C. B., and Gibbs, J. L. (2006). Unpacking the Concept of Virtuality: The Effects of Geographic Dispersion, Electronic Dependence, Dynamic Structure, and National Diversity on Team Innovation. *Administrative Science Quarterly*, 51(3), 451–495. doi: 10.2189/asqu.51.3.451

Gill, R., and Pratt, A. (2008). In the social factory? Immaterial labour, precariousness and cultural work. *Theory, Culture and Society*, *25*(7–8), 1–30.

Gilson, L. L., Maynard, M. T., Young, N. C. J., Vartiainen, M., and Hakonen, M. (2015). Virtual Teams Research:10 Years, 10 Themes, and 10 Opportunities. *Journal of Management*, 41(5), 1313–1337. doi: 10.1177/0149206314559946

Gregg, M. (2011). *Work's intimacy*. Cambridge, MA: Polity Press Cambridge.

Hertel, G., Geister, S., and Konradt, U. (2005). Managing virtual teams: A review of current empirical research. *Human Resource Management Review*, 15(1), 69–95. doi: https://doi.org/10.1016/j.hrmr.2005.01.002

Hesmondhalgh, D., and Baker, S. (2011). Toward a political economy of labor in the media industries. *The Handbook of Political Economy of Communications*, 381–400.

Hilton, M. (2008). Skills for Work in the 21st Century: What Does the Research Tell Us? *Academy of Management Perspectives*, 22(4), 63–78.

Hoch, J. E., and Kozlowski, S. W. J. (2014). Leading virtual teams: Hierarchical leadership, structural supports, and shared team leadership. *Journal of Applied Psychology*, 99(3), 390–403. doi:10.1037/a0030264

Hollister, M. (2011). Employment stability in the US labor market: Rhetoric versus reality. *Annual Review of Sociology*, 37, 305–324.

Horowitz, S. and Rosati, F. (2014). Freelancing in America: A national survey of the new workforce. Freelancers Union and Elance-oDesk. Accessed on: https://www. freelancersunion.org/blog/dispatches/2014/09/04/53million, 1(10), 2015.

Horton, J. and Chilton, L. (2010). The labor economics of paid crowdsourcing. *Proceedings of the 11th ACM Conference on Electronic Commerce*, 209–218.

Howe, J. (2009). *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. London: Random House Business Books

Huang, R., Kahai, S., and Jestice, R. (2010). The contingent effects of leadership on team collaboration in virtual teams. *Computers in Human Behavior*, 26(5), 1098–1110. doi: https://doi.org/10.1016/j.chb.2010.03.014

Huws U. (2017) Where Did Online Platforms Come From? The Virtualization of Work Organization and the New Policy Challenges it Raises. In: Meil P., Kirov V. (eds) *Policy Implications of Virtual Work. Dynamics of Virtual Work*. Palgrave Macmillan, Cham.

Irani, L. (2015). The cultural work of microwork. *New Media and Society*, 17(5), 720–739. doi: 10.1177/1461444813511926

Jenkins, H. (2006). *Confronting the challenges of participatory culture? Media Education for the 21 century*. MacArthur Foundation Report.

Jung, A. K. (2015). Interpersonal and societal mattering in work: A review and critique. *The Career Development Quarterly*, 63(3): 194–208.

Kalekin-Fishman, D., and Langman, L. (2015). Alienation: the critique that refuses to disappear. *Current Sociology*, 63(6): 916–933.

Kelliher, C. and Anderson, D. (2010). Doing more with less? Flexible working practices and the intensification of work. *Human Relations* 63(1), 83–106.

Kingsley, S. C., Gray, M. L., and Suri, S. (2015). Accounting for Market Frictions and Power Asymmetries in Online Labor Markets. *Policy and Internet*, *7*(4), 383–400.

Kirkman, B. L., Rosen, B., Tesluk, P. E., and Gibson, C. B. (2004). The Impact of Team Empowerment on Virtual Team Performance: The Moderating Role of Face-to-Face Interaction. *The Academy of Management Journal*, 47(2), 175–192.

Kittur, A., Chi, E. H., and Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 453–456.

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Horton, J. (2013). The future of crowd work. Paper presented at the Computer supported cooperative work 2013, San Antonio, Texas, USA.

Kneese, T., and Rosenblat, A. (2014). Understanding Fair Labor Practices in a Networked Age. *Open Society Foundations' Future of Work Commissioned Research Papers.*

Kolbæk, R. (2015). Nursing students' attitudes towards ICT in education and clinic in Denmark. *International Journal of Media, Technology and Lifelong Learning*, 11(1), 1–14.

Kosonen, M., Gan, C., Vanhala, M., and Blomqvist, K. (2014). User Motivation And Knowledge Sharing In Idea Crowdsourcing. *International Journal of Innovation Management*, 18(5), 1450031–1.

Kücklich, J. R. (2009). Virtual worlds and their discontents precarious sovereignty, governmentality, and the ideology of play. *Games and Culture*, *4*(4), 340–352.

Kuehn, K., and Corrigan, T. F. (2013). Hope labor: The role of employment prospects in online social production. *The Political Economy of the Media, 1(1),* 9–25.

Kuek, S. C., Paradi-Guilford, C., Fayomi, T., Imaizumi, S., Ipeirotis, P., Pina, P., and Singh, M. (2015). *The global opportunity online outsourcing*. World Bank Group.

Leftheriotis, I., and Giannakos, M. N. (2014). Using social media for work: Losing your time or improving your work? *Computers in Human Behavior*, 31(Supplement C), 134–142. doi: https://doi.org/10.1016/j.chb.2013.10.016

Lehdonvirta, V. and Ernkvist, M. (2011). *Knowledge Map of the Virtual Economy*. Washington, DC: World Bank.

Leonardi, P. M. (2015). Ambient Awareness and Knowledge Acquisition: Using Social Media to Learn "Who Knows What" and "Who Knows Whom". *Mis Quarterly*, 39(4), 747–762.

Leonardi, P. M., and Treem, J. W. (2012). Knowledge management technology as a stage for strategic self-presentation: Implications for knowledge sharing in organizations. *Information and Organization*, 22(1), 37–59. doi: https://doi.org/10.1016/j.infoandorg.2011.10.003

Lewis, W., Agarwal, R., and Sambamurthy, V. (2003). Sources of Influence on Beliefs about Information Technology Use: An Empirical Study of Knowledge Workers. *Mis Quarterly*, 27(4), 657–678.

Liao, C. (2017). Leadership in virtual teams: A multilevel perspective. *Human Resource Management Review,* 27(4), 648–659. doi: https://doi.org/10.1016/j.hrmr.2016.12.010

Manyika, J., Bughin, J., Lund, S., Nottebohm, O., Poulter, D., Jauch, S., and Ramaswamy, S. (2014). *Global flows in a digital age: How trade, finance, people, and data connect the world economy*. McKinsey Global Institute.

Martin, G., Parry, E., and Flowers, P. (2015). Do social media enhance constructive employee voice all of the time or just some of the time? *Human Resource Management Journal*, 25(4), 541–562. doi: 10.1111/1748-8583.12081

Martins, L. L., Gilson, L. L., and Maynard, M. T. (2004). Virtual Teams: What Do We Know and Where Do We Go From Here? *Journal of Management*, 30(6), 805–835. doi: http://dx.doi.org/10.1016/j.jm.2004.05.002

Mesmer-Magnus, J. R., DeChurch, L. A., Jimenez-Rodriguez, M., Wildman, J., and Shuffler, M. (2011). A meta-analytic investigation of virtuality and information sharing in teams. *Organizational Behavior and Human Decision Processes*, 115(2), 214–225. doi: http://dx.doi.org/10.1016/j.obhdp.2011.03.002

McKercher, C., and Mosco, V. (2008). *The laboring of communication: Will knowledge workers of the word unite*. Lanham, MD: Lexington Books

Nakatsu, R. T., Grossman, E. B., and Iacovou, C. L. (2014). A taxonomy of crowdsourcing based on task complexity. *Journal of Information Science*, 40(6), 823–834.

Orlikowski, W. J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, 3(3), 398–427. doi: 10.1287/orsc.3.3.398

Ortiz de Guinea, A., Webster, J., and Staples, D. S. (2012). A meta-analysis of the consequences of virtualness on team functioning. *Information and Management*, 49(6), 301–308. doi: http://dx.doi.org/10.1016/j.im.2012.08.003

Pasmore, W., Francis, C., Haldeman, J., and Shani, A. (1982). Sociotechnical systems: A North American reflection on empirical studies of the seventies. *Human Relations*, 35(12), 1179–1204. doi: 10.1177/001872678203501207

Piezunka, H., and Ander, L. D. (2015). Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. *Academy of Management Journal*, 58(3), 856–880.

Pinch, T. J., and Bijker, W. E. (2012). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. In W. E. Bijker, T. P. Hughes, T. J. Pinch, and D. G. Douglas (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology*. Cambridge, MA: The MIT Press.

Postigo, H. (2016). The socio-technical architecture of digital labor: Converting play into YouTube money. *New Media and Society*, 18(2), 332–349.

Powell, W. W. (1987). Review essay: Explaining technological change. *American Journal of Sociology*, 93(1), 185–197.

Pratt, A. C., and Gill, R. C. (2000). *New media centre user-networks.* London: Arts Council of England.

Rainie, L. and Wellman, B. (2012). *Networked: The New Social Operating System*. Cambridge, MA: MIT Press.

Rosenblat, A. and Stark, L. (2015). *Uber's Drivers: Information Asymmetries and Control in Dynamic Work.* Available at SSRN 2686227.

Ruggieri, A., Mosconi, E. M., Poponi, S., and Braccini, A. M. (2015). Strategies and policies to avoid digital divide: the Italian case in the European landscape. In *From Information to Smart Society*, 221–230. Springer International Publishing.

Sarker, S., Ajuja, M., Sarker, S., and Kirkeby, S. (2011). The role of communication and trust in global virtual teams: A social network perspective. *Journal of Management Information Systems*, 28(1), 273–309.

Scholz, T. (ed.) (2013). *Digital Labor: The Internet as Playground and Factory*. New York: Routledge.

Silberman, M., Ross, J., Irani, L., Zaldivar, A., and Tomlinson, B. (2010). Who are the crowdworkers?: shifting demographics in mechanical turk. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems,* 2863–2872.

Smith, A. (2016). *Shared, Collaborative and On Demand: The New Digital Economy*. Pew Research Center Report.

Spreitzer, G., Cameron, L., and Garrett, L. (2017). Alternative work arrangements: Two images of the new world of work. *Annual Review of Psychology and Organizational Behavior*, 4, 473–499.

Tapscott, D., and Williams, A. D. (2008). *Wikinomics: How mass collaboration changes everything.* New York: Penguin.

Terranova, T. (2004). *Network culture: Politics for the information age.* London: Pluto Press.

Van Dijck, J. and Nieborg, D. (2009). Wikinomics and its discontents: a critical analysis of Web 2.0 business manifestos. *New Media and Society*, *11*(5), 855–874.

Venkatesh, V., and Goyal, S. (2010). Expectation disconfirmation and technology adoption: Polynomial modelling and response surface analysis. *Mis Quarterly*, 34(2), 281–303.

Wolf, M. (2016). Seven ways technology has changed us. *Financial Times*.

Zwick, D., Bonsu, S. K., and Darmody, A. (2008). Putting Consumers to Work: Co-creation and new marketing govern-mentality. *Journal of Consumer Culture*, 8(2), 163–196.

DOI: 10.18261/9788215031583-2018-05

Chapter 4
Building and Managing Reputation: Current Debates and Future Directions

PEGGY SIMCIC BRØNN AND ALEXANDER BUHMANN

ABSTRACT This chapter takes up the subject of reputation and its strategic importance for organizations. We provide an overview of generally accepted definitions of reputation and recognize the complexity of reputation by introducing a discussion on why managing reputation is a wicked problem and how organizations can best "solve" it by building awareness of reputation into organizational DNA. The chapter offers insight into a number of areas where future research might better assist all organizations in realizing the potential of their reputation.

KEY WORDS: reputation | systems approach | stakeholder | intangible asset

4.1 INTRODUCTION

In 2004, scholars at BI Norwegian Business School were instrumental in introducing the first research-based reputation measurement instrument to Norway, the Reputation Institute's Reputation Quotient, now called RepTrak (Gardberg and Fombrum, 2002). Prior to 2004, the word reputation was rarely mentioned in the Norwegian business press. In the following five years, reputation gained a great deal of attention as the number of articles in the media mentioning reputation increased by a factor of eight, from slightly over 1,000 to more than 8,000 in 2009 (A-tekst, 2018).

Reputation has now become a buzzword; so-called reputation experts have been made a laughing stock by the media, and it has been proposed that leaders are tired of hearing their communication people talk about reputation (Brønn and Ihlen, 2009). This flies in the face of the fact that reputation remains one of the most important assets of modern organizations. It assists in building competitive advan-

tage because it is difficult to imitate, acquire, or replace; it builds legitimacy, making organizations more resilient to crises; and it increases their room to maneuver in organizational change and innovation. The dismissive attitude toward reputation also ignores its complexity. Everything an organization does contributes to either a good or a bad reputation, and building a good reputation over time requires engaging the entire organization and ensuring that the performance of everyone in the organization is of high quality. This is relevant for the private and public sectors, including non-profit organizations and governmental agencies.

That reputation is of interest not only to organizations, but also to society in general, is evident in the scope and number of reputation rankings worldwide. Some estimates indicate that there are more than 500 company rankings published annually, ranging from best place to work for minorities or for women, most admired companies, most environmentally friendly, most ethical, and so on. The release of these surveys is met with much fanfare and many organizations use their high rankings to generate publicity and to build their brands. For the losers, the results may result in panic and in possible loss of reputation, subsequent financial catastrophe and increasing pressure for accountability (Busuioc and Lodge, 2016). Regardless, organizations will spend resources in the form of time, money and knowledge to maintain or improve reputation, or to build it up.

It is impossible to adequately cover the extensive research on reputation in a single chapter; however, we provide an overview of generally accepted definitions of reputation and recognize its complexity by introducing a discussion on why managing reputation is a wicked problem and how organizations can best "solve" it by building awareness of reputation into organizational DNA. The chapter concludes by offering insights into a number of areas where future research might better assist all organizations in realizing the potential of their reputation, not just in the marketplace, but also, most importantly, in society in general.

4.2 DEFINING REPUTATION

Reputation has been described as a broad "portmanteau concept" with many interpretations (Brønn and Brønn, 2015). Research has uncovered anywhere from 16 to 50 different definitions (Bennett and Kottasz, 2000; Barlett et al., 2006; Dowling, 2016). Furthermore, the various academic disciplines differ in their views of reputation. For economists, reputation is the sum of those characteristics or signals that describe a firm's possible behavior in special situations. Accounting sees reputation as one of several types of intangible assets that are difficult to measure, but that create long-term value. For marketing, reputation comprises the associations

that individuals have with an organization's name, while in the field of communication, reputation is defined as the organizational characteristics that develop from the relationships the organization has with its environment. In organizational theory, reputation is seen as the cognitive representation of organizations as stakeholders acquire meanings of the organization, and in sociology it is a social construction that results from the relationships that the organization establishes with its stakeholders in their common institutional environments: a good reputation is an indicator of legitimacy (van Riel and Fombrun, 2007).

Barnett et al. (2006) group the definitions into three broad categories:

- *Awareness*: Reputation as the attention that a stakeholder gives an organization, but not necessarily implying a judgment—reputation is seen merely as a perception or impression.
- *Assessment*: Reputation as a judgment, an estimate, an evaluation or a gauge—reputation says something about the status of an organization.
- *Asset*: Reputation as something of value and significance to the organization—reputation is a resource, an intangible, financial, or economic asset.

Awareness is a key concept, as we know that people's perceptions of an organization's performance are based on communication about and from the organization, their own experience with the organization, and what others tell them about the organization. Dowling (2016) refers to Lange et al. (2011), who identify three dominant conceptualizations of corporate reputation. Organizations need:

- to be well known—the salience or prominence of the organization;
- to be known for something—beliefs about an organization's distinctive characteristics and/or behaviors; and
- to have a generalized favorability—an overall evaluation of being good or attractive.

Being known is a prerequisite for the other two.

One of the most cited definitions of reputation is offered by Fombrun (1996), who sees reputation as "a perceptual representation of a company's past actions and future prospects that describes the firm's overall appeal to all of its key constituents when compared with other leading rivals" (p. 72). An equally relevant definition is provided by Barnett et al. (2006), who defined reputation as observers' collective assessment of a business based on the perceptions of the financial, social and environmental consequences a business has over time. This definition thus emphasizes

how reputation is a value that someone *outside* the business determines. At the same time, the definition states that this is a socially created value.

An enterprise's reputation will therefore be influenced by peoples' opinion based on the direct experience they have had with products, behavior, character and the like; what they are told by others; and by organization's past behavior. This knowledge is influenced by the individual values of each stakeholder or stakeholder group. Over time, all of these impressions create a reputation capital—the intangible resource that should strengthen the organization's competitive advantage or, in the public sector, its status among the general population.

The definition above also helps to draw attention to the *evaluation* of an organization. Reputation is the appeal an organization has in its environment, and it is being assessed in comparison with other organizations. At the same time, it is worth noting that the definition emphasizes that reputation has a time dimension—a reputation is built over time and should last over time.

As a construct of judgment (assessment) reputation comprises both affective (emotional) and cognitive components (Ingenhoff and Buhmann, 2016). Reputation as an emotional aspect is defined by Fombrun and van Riel (2004) as the degree to which people admire and respect the organization, trust the organization, feel good about the organization or think the organization has an overall good reputation. Cognitive components of the construct are represented by people's belief in or judgment about an organization's performance on seven common primary dimensions: (1) leadership, (2) products and services, (3) financial performance, (4) innovation, (5) workplace environment, (6) governance, and (7) citizenship. These are proxy variables directly associated with measurements of quality where reputation management consists of ensuring superior performance. The relationship between the affective and the cognitive aspects of reputation are shown in Figure 4.1 from Reputation Institute. The emotional dimension is captured by the RepTrak Pulse.

There is also a significant literature reflecting the considerable empirical attention to understanding how reputation impacts supportive behavior toward an organization. This includes willingness to purchase products, invest in a company, recommend company/products, work for a company, or give them the benefit of the doubt in crises (see for example, Money et al., 2016).

It is important to note that research on reputation is currently broadening considerably, going beyond the common focus on organizations. Recent work on the reputation and image of countries is a good example of this. The role of the country image, how it is created and its effects are of major interest for those working in international relations, international marketing, politics, trade, or tourism (Buhmann and Ingen-

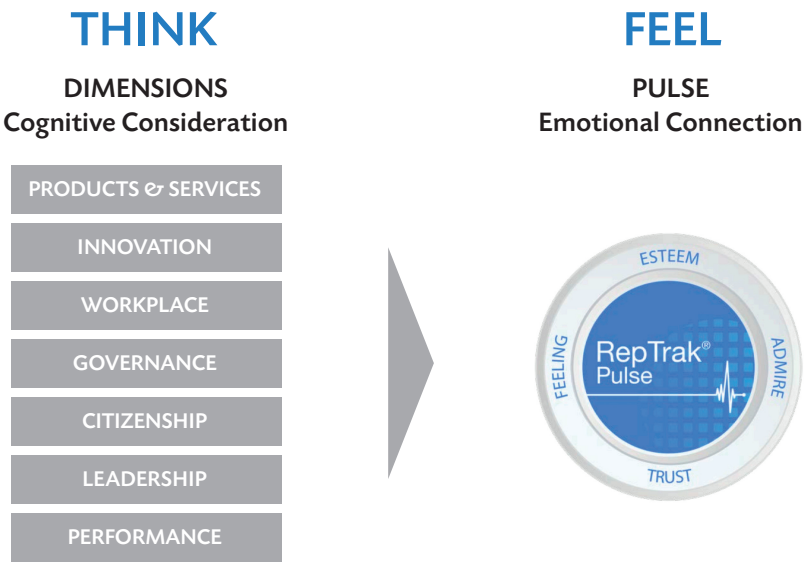

FIGURE 4.1

hoff, 2015a). A country's image and reputation is becoming more important and its impact is seen through effects on the level of exports and foreign direct investments, the stability of international relations, the prosperity of national tourist industries, the attractiveness of domestic labor markets and education systems, or the degree of a country's political and economic influence in the international system.

Recent research (e.g., Buhmann and Ingenhoff, 2015b; Buhmann 2016a, 2016b) has resulted in new integrative models where statistical analyses can show how functional, normative, and aesthetic beliefs about a country affect the formation of the emotional country image dimension—showing the country's "ability to attract." Furthermore, we can see how the emotional dimension of the country image mediates the effect of the cognitive dimensions on people's behavior. Such behavioral effects can be analyzed regarding a wide variety of outcome variables such as the willingness to politically support a country, invest in a country, or travel to a country.

It is also possible to study the effects of corporate crises on country reputation. A recent study by Ingenhoff, Buhmann, White, Thang and Kiousis (2018), found that how the media reports on crises involving "nationally branded" corporations affects the perception of their home country. An example from Norway illustrates the interplay between national corporate brands and country image. In 2017, Norway had an exceptional image, ranking number one in the world for happiness by Sustainable Development Solutions Networks, and sixth on the Reputation Insti-

tute's ranking of the world's most reputable countries. When Statoil's CEO was pictured in the media sitting in an all-male meeting with Donald Trump at the 2018 World Economic Forum in Davos, there was negative pushback in Norway from influential personalities on Facebook. That "optic" crashes with the international audience's expectations regarding Norway's position of valuing gender parity, and the assumption that its business leaders would feel the same and act accordingly. There is no evidence that the discussion was greater than in Norway (some media did note the fact they were all men), but the danger of it snowballing certainly existed, particularly when gender parity is a recurring theme at the World Economic Forum. This could have raised questions not only about Statoil's behavior, but also about Norway's.

Organizations may believe there is a quick fix to reputation, but a quick look at the list of drivers of reputation (product and services, workplace, leadership, performance, citizenship, governance and innovation) puts the lie to this assumption. There is no one person within an organization in charge of all of these drivers, who is capable of managing all of the stakeholders associated with each driver, who is familiar with the diverse measurements of success, and so on. When everything that an organization does can theoretically impact its reputation, it is clear that reputation-building is not a function that can be left to one individual or to one department alone. It becomes the responsibility of everyone in an organization. Because an organization's reputation can act as a standard governing behavior, Balmer (2003) argues that, *everyone* should be encouraged to ask the question: "Would my actions be in line with the company's good/bad reputation?" This is reflected in Balmer's D.E.A.R. principle: Decisions, Evaluated, Against the Reputation.

The discussion above underscores the complexity of reputation: It has many interpretations; it is a judgment held by a diversity of stakeholders, all of whom have diverse and often conflicting views and priorities; it is something that in the long-run will never be perfect; and it can be influenced positively or negatively at any time by both internal and external actors. The notion of reputation and reputation management thus bears many of the characteristics of a class of problems called wicked problems (Rittel and Webber 1973). Wicked problems are not evil or "cool" problems, but rather problems that are resistant to long-lasting solutions, that are hard to grasp and change when efforts are made to deal with them, and that have complex roots and diverse stakeholders.

There are no specific solutions to wicked problems like reputation, but it is possible to deal with them. One strategy for dealing with wicked problems is to adopt a learning orientation to engage the entire organization, in other words thinking about reputation needs to be built into the very DNA of an organization where all

"decisions are evaluated against reputation". One approach is explained in the following section.

4.3 REPUTATION: YOU'VE GOT TO BUILD IT IN

Dowling (2016) argues that reputation must be "built in" to an organization's DNA by coupling reputation-building actions closely to the organization's strategy. This approach is perceived as a more natural component of the organization's activities and can be expected to give positive associations with the organization in the minds of a broader range of stakeholders. This is in contrast to a bolted-on approach where efforts appear to be an afterthought or an add-on; the activities seem to be insincere to external stakeholders, and thus lower the estimation of the firm in their minds.

A bolted-on reputation management approach would be to apply a quick and dirty solution to some more fundamental problem that is, for example, generating consistently low reputation ratings. It is not unusual to hear of even large international firms running campaigns such as "Number 1 in 2001" after a poor performance on a reputation ranking the previous year. The bolted-on strategy may have some short-term benefits where improving stakeholders' immediate perceptions of the firm results in an improvement in the ranking scores and eases the immediate problem. Over time, however, the fundamental problem will reassert itself.

The built-in approach is based on implementing a more fundamental solution that will also ease the problem's symptoms, but will do so through other means and with some time delay. When managers are alerted to the potential unintended consequences of applying short-term and immediate solutions, they will hopefully look deeper at underlying causes and not patch the problem. In reality, a balanced approach to solving hard problems is required. In this view, the strategy that is most likely to succeed is one that attempts to find a balance between short- and long-term actions. Moreover, what works for one organization may not work for another. Making this happen requires focusing attention on the total organizational system instead of just the individual parts or properties of the parts; not just product, not just the financial, not just workplace, and not just leadership.

Because organizational reputation reflects an assessment of performance across all functions and hierarchical levels, it is not possible to reduce the reputation construct directly back to the component actions that are carried out by the many organizational actors. Reputation is neither predictable nor deducible from the functioning of lower-level organizational components. Organizations are complex systems, and initiatives intended to improve on one dimension of reputation will affect the other dimensions to varying degrees. As such, the organizational reac-

tion may be counter-intuitive with the net result being an overall decrease in key performance indicators, and a loss of reputation.

Further complicating the situation is that stakeholders often have different meanings and expectations for what constitutes good behavior for the various drivers of reputation. As firms strive to meet stakeholder demands and expectations, gaps can occur between various organizational members' understanding of what these expectations are and what behavior organizations should deliver to meet the expectations. Consequently, this requires organizations to minimize gaps between what is expected of them and the behavior delivered by them. This also reduces reputation risk (see Brønn, 2012). The challenge for any organization is to build in a mental model, if you will, that gets everyone in the organization pulling together to minimize quality gaps. This organizational mindset is not easy to achieve.

The following discussion outlines some areas where more or continued research can hopefully shed light on some of the complexities of reputation and reputation management.

4.4 A REPUTATION RESEARCH AGENDA

Even though there is an impressive literature on reputation, on what it is and isn't, what it can do, its impact, and so on, there are still many areas of inquiry that, if explored, can help organizations build, maintain or rebuild reputation, and not the least help them create conditions for a built-in systemic approach to reputation. For example, most agree with the axiom that it takes a long time to build reputation but almost no time at all to destroy it. This is often followed by the question: how long does it take to rebuild? This has been explored to a minor extent (see for example, Gaines Ross, 2008; deHaan, 2017) but is still of great interest across industries, countries, and organizations.

Another area of interest is the impact of new digital technologies on an organization's reputation. This can refer to new forms of communication and interaction that emerge with new digital technologies, such as in rating portals or social networks. An important domain in this regard is the influence on an organization's reputation from so-called uncurated third-party rankings, such as Yelp or TripAdvisor, where the third parties are anonymous members of the general public. Similar to rankings such as RepTrak, these polls can either boost an organization's reputation, or have devastatingly negative effects. There is growing research on how organizations can take a more proactive role in online communication, including adding their own voice to the conversation (see for example, Aula, 2011).

The impact of new digital technologies can also refer to reputational implications of the new technologies themselves. A critical aspect of digitalization is the growing use of complex algorithm-based systems by organizations to interpret and predict consumer behavior, support and carry out operations, or drive recommendation and filtering systems. While organizations today make extensive use of complex machine learning algorithms, hardly anyone is able to fully account for their workings. Current research at BI's Centre for Corporate Communication and the Nordic Centre for Internet and Society adds to knowledge in this area by: a) mapping common kinds of reputational threats caused by algorithms; b) reviewing in particular the inherent opacity *threat* of algorithms, which arises from poor transparency and leads to distinct challenges for accountability; c) suggesting normative principles to assess algorithmic accountability based on a discourse-ethical framework (i.e.: access to deliberation, access to information, inclusion of all arguments, and responsiveness); and d) applying these principles to cases where algorithms become major threats to organizational reputation (Buhmann, Passmann, and Fieseler, forthcoming).

The reputation of public sector entities is of growing concern and more research is needed on the drivers of reputation in the public sector. This is important, as it is generally understood that the public sector is a special case and that reputation-building in the public sector cannot be approached in the same manner as in the private sector. Bottom line performance in the public sector is not about profits, but rather benefits to the community. While there is a substantial amount of research on service satisfaction in the public sector (see for example, Rowley, 1998; Heintzman and Marsden, 2005), there is little research on reputation and the public sector. According to Sørensen (2009), people today have higher expectations of the public sector, are more skeptical toward experts and public authority, and are less confident that these people are working for their or society's interests. Additionally, people want influence over and freedom to choose who delivers public service. They want more and better information, particularly about who can deliver the best quality.

Of continued interest will be the pressing question of measurement and evaluation, not only of reputation but also for those activities recognized as building reputation, particularly communication efforts (Buhmann, Likely, and Geddes, 2018). Key for organizations is realizing that the use of one-size-fits-all reputation measurements will not necessarily help them with their stakeholders, in their market, or in their operating environment. It is essential to find a valid and reliable measurement tool that is the best and most relevant tool for their own organization, which in many cases will mean creating their own.

Some firms in highly competitive environments conduct monthly customer surveys, but no reputation surveys. Some organizations do nothing, and some have bi-annual or annual reviews. What is important is to recognize that some measurement must be done; not only to create accountability for the many activities that build reputation, but also to generate insights and learning on the inevitable changes in the dynamic relationship of the organization and its environment. Reputation is an asset that has financial value, and the only way to know what stakeholders' expectations are—and if they are being met—is to ask them. Ignoring reputation measurement is not an option.

Reputation as a component of an organization's overall risk assessment and management process is also an important emerging research area. By recognizing reputation as a component of risk, i.e., considering stakeholder expectations in decision-making, organizations signal that they are concerned about their role in society and their willingness to include a broader range of perspectives and voices into decision processes. This is a complex undertaking, however, and organizations need better guidelines on how to engage responsibly and productively. This is important not only for big, visible organizations, but also for smaller ones who are unable to absorb the loss of key customers, a supplier default, a lawsuit or a credit problem, and where reputation damage can mean the end of business.

Paradoxically, damage to reputation today is most likely to come from a friendly source: an organization's own management. This often happens through poor or unethical decision-making that creates circumstances that could have been avoided. For example, it is estimated that Volkswagen's (VW) managers' decision to adjust their cars' emission controls to make the cars appear "greener" resulted in a decline in market capitalization of about 20%, or over €20 billion (Investopedia, 2016). Part of the decline is a result of damage to VW's reputation as a green company; people stopped believing what the firm was saying. In January 2017, VW experienced another management-made crisis as the firm had to admit it used monkeys and humans to test exhaust systems. In both cases, one of the first persons to be fired was the communication executive, even though they were not directly involved in the decision process. Arguably, this connotes a culture that is not transparent and where those on the front line appear to be the scapegoats for hidden and closed decision-making. The question is whether this behavior is sustainable in a technology-mediated environment. This type of crisis is not isolated to the private sector. Recent bad behavior by leaders and managers has emerged in political parties in Norway and at one of the biggest international non-profit organizations in the UK.

4.4.1 CHANGING HOW REPUTATION IS STUDIED: FROM OUTCOMES TO PROCESS

The fall of organizations' reputations strengthens the argument for changing the way reputation is studied. The internal breakdown of management control, the "wickedness" of the reputation problem and the need for a built-in approach, illustrate the need for research employing a systems thinking approach to building reputation, with its stronger focus on process as opposed to outcome (see for example, Kraatz and Love, 2006). What occurs within an organization that leads to it being perceived in certain ways is much more difficult to evaluate than gathering results from a reputation survey; it requires exploring reflective learning, the learning skills associated with systems thinking. These skills are intuitive and easy to understand at the intellectual level, but are notoriously difficult to implement. This difficulty comes from the need to change the fundamental perspective of understanding the world. This includes overcoming the ingrained habit of focusing on specific events and cultivating a habit of stepping back to look at patterns of events over time. This means moving from a reactionary thinking mode to a generative thinking mode. In the latter, decision makers attempt to develop endogenous or internal as opposed to externally-attributed explanations and solutions to the observed problematic behaviors. For example, looking inside for why reputation is declining, not blaming it on external causes.

The transition to systems thinking is also hampered by organizational and social cultures, both of which are strongly biased to considering only the short-term and relying on linear cause and effect processes. By definition, reputation building is a long-term process, making short-term thinking inappropriate.

Advancing the systems thinking approach to reputation, with its emphasis on process, requires research in two areas. First, exploring and evaluating the role of management in fostering a learning environment. Second, studying the overall learning orientation of the organizations, which is important because learning organizations are more effective in dealing with wicked problems. This also addresses the organization's culture for learning (see for example, Nevis, DiBella, and Gould, 1995). Together, an assessment of managerial practices across functions and of organizational culture with respect to learning can identify opportunities for improving organizational learning capabilities, thereby supporting reputation management activities. For example, firms that score high on the RepTrak survey tend to score high on all of the seven drivers of reputation. It would be an important contribution to analyze these firms to see to what extent their learning characteristics are consistent with the challenges associated with managing their reputation. This research would help organizations to understand the basis of their

particular reputation and provide specific guidance on how to manage their behavior more effectively and, consequently, its impact on their reputations. These insights would be invaluable not only for the private sector, but for all organizations, and even countries.

4.5 CONCLUSION

Because reputation is an indirect intangible resource, the role reputation plays in contributing to organizational performance is more ambiguous and, consequently, easy to ignore or to treat only superficially. On the surface, reputation is a concept that people feel comfortable with; at least, they have an intuitive understanding of what it is and its value. They are comfortable with its everyday use. However, effective management of a dynamic and "wicked" resource such as reputation requires a high level of insight. Events and patterns that build or harm reputation do not just happen; there is something that causes the event or pattern, and that maintains the behavior over time.

Reputation is likely one of the most difficult strategic issues that managers confront (Brønn and Brønn, 2017). The procedures that go into making a firm's products and assuring quality are relatively clearly understood and possible to control. Product quality and financial performance can be improved, innovative processes can be put into place, and decent and fair wages can be paid to employees. Reputation, on the other hand, is an emergent characteristic of the organization as a whole and, as such, cannot be "managed" directly.

There is some speculation that the study and measurement of reputation is at a standstill and that new insights are necessary if organizations are to realize the full worth of this valuable intangible asset. It is thus imperative that we continue research into reputation and its related fields. Organizations need better ways of understanding reputation, its formation, and its impact on their own type of organization (or country) along with more relevant tools for helping them instill an organization-wide systems approach to building reputation.

The prominent economist John Kay (1993) was clear when he stated that as a distinctive capability that accrues competitive advantage to an organization, reputation is an organization's most important commercial mechanism for conveying information about itself. There is no doubt that a good reputation is a resource that gives competitive advantage because it is difficult for competitors to imitate it, acquire it, or replace it. It is a resource that should be protected, just as an organization protects its other resources.

REFERENCES

Aula, P. (2011). Meshworked reputation: Publicists' views on the reputational impacts of online communication, *Public Relations Review*, 37(1), 28–36.

Barnett, M.L., Jermier, L.M. and B.A. Lafferty (2006). Corporate reputation: The definitional landscape, *Corporate Reputation Review*, 9(1), 26–38.

Bennett, R. and R. Kottasz (2000). Practitioner perceptions of corporate reputation: An empirical investigation, *Corporate Communications: An International Journal*, 5(4), 224–234.

Brønn, P. S. (2012). Adapting the PZB Service Quality Model to Reputation Risk Analysis and the Implications for CSR Communication, *Journal of Communication Management,* 16, 1, 77–94.

Brønn, C. (2016). Applying Systems Thinking to Reputation. In (C. Carroll ed.) *Sage Encyclopedia of Corporate Reputation.* Sage: Thousand Oaks, CA.

Brønn, P. S. and C. Brønn (2017). Systems Thinking: A Method for Reducing Reputation Risk. *International Studies of Management and Organisation* 2017; Volume 47 (3).

Brønn, C. and P. S. Brønn (2015). A Systems Approach to Understanding How Reputation Contributes to Competitive Advantage. *Corporate Reputation Review*. 18.(2), 69–86.

Buhmann, A. (2016a). *Measuring Country Image: Theory, Method, and Effects*. Springer VS: Wiesbaden.

Buhmann, A. (2016b). The constitution and effects of country images: Theory and measurement of a central target construct in international public relations and public diplomacy. *Studies in Communication Sciences*, 16(2) 182–198.

Buhmann, A. and D. Ingenhoff (2015a). The 4D-Model of the country image. An integrative approach from the perspective of communication management, *International Communication Gazette*, 77(1), 102–124.

Buhmann, A. and D. Ingenhoff (2015b): Advancing the country image construct from a public relations perspective: from model to measurement, *Journal of Communication Management*, 19(1), 62–80.

Buhmann, A., Likely, F. and D. Geddes (2018). Communication Evaluation and Measurement: Connecting Research to Practice. *Journal of Communication Management*, 22(1), 113–119.

Busuioc, M. and M. Lodge (2016). Reputation and Accountability Relationships: Managing Accountability Expectation through Reputation. *Public Administration Review*, 77(1), 99–100.

deHaan, E. (2017) The Financial Crisis and Corporate Credit Ratings. *The Accounting Review*, Vol. 92, No. 4, 161–189.

Dowling, G. (2006) Reputation risk: It is the board's ultimate responsibility, *Journal of Business Strategy* 27(1), s. 59–68.

Dowling, G. (2016). Defining and Measuring Corporate Reputations. *European Management Review*, Vol. 13, 207–223.

Dowling, G. (2016). *Winning the Reputation Game: Creating Stakeholder Value and Competitive Advantage*, MIT Press.

Fombrun, C. J .and C.B.M. van Riel, (1997). The reputational landscape. *Corporate Reputation Review,* 1 (1/2), 5–13.

Fombrun, C.J. and M. Shanley (1990) What's in a Name? Reputation Building and Corporate Strategy, *Academy of Management Journal,* 33(2), s. 233–258.

Gaines Ross, L. (2008). *Corporate Reputation: 12 Steps to Safeguarding and Recovering Reputation.* John Wiley & Sons: Hoboken, NJ.

Gardberg, N.A. and C.J. Fombrun 2002. The Global Reputation Quotient Project: First Steps towards a Cross-Nationally Valid Measure of Corporate Reputation, *Corporate Reputation Review* 4(4), 303–308.

Heintzman, R. and B. Marson (2005). People, service and trust: is there a public sector service value chain?, *International Review of Administrative Sciences*, 71, 549–575.

Ingenhoff, D., Buhmann, A., White, C., Zhang, T and S. Kiousis (2018). Reputation Spillover: Corporate Crises' Effects on Country Reputation. *Journal of Communication Management*, 22(1), 96–112.

Ingenhoff, D. and A. Buhmann (2016). Advancing PR measurement and evaluation: Demonstrating the properties and assessment of variance-based structural equation models using an example study on corporate reputation, *Public Relations Review*, Vol. 42(3), 418–431.

Investopedia, (2016). VW Scandal: How Has It Impacted Volkswagen's Stock? (VLKAY), 15 October, https://www.investopedia.com/news/vw-scandal-how-has-it-impacted-volkswagens-stock-vlkay/

Kay, J. (1993). *Foundations of corporate success*. New York: Oxford University Press.

Kraatz, M.S. and E. G. Love (2006). Studying the dynamics of reputation—A framework for research on the reputational consequences of corporate actions. In: (B. Boyd, T. R. Crook, J. Lê and A. Smith eds.), *Research Methodology in Strategy and Management,* Vol. 3, 343–383. Emerald Books:

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

Nevis, E. C., DiBella, A. J. and J. M. Gould (1995). Understanding organizations as learning systems, *Sloan Management Review*, Winter, 73–85.

Rittel, H.W.J., and M.M. Webber (1973). Dilemmas in a General Theory of Planning, *Policy Sciences* 4, 155–169.

Rowley, J. (1998). Quality measurement in the public sector: Some perspectives from the service quality literature, *Total Quality Management*, 9, 2/3, 321–333.

Sørensen, R. J. (2009). *En effektiv offentlig sektor*, Universitetsforlaget: Oslo, Norway.

DOI: 10.18261/9788215031583-2018-06

Chapter 5
Future of Branding in the Digital Age

LARS ERLING OLSEN

ABSTRACT Digitalization of markets, media channels, and consumers' decision making challenges brand management. A key question is whether the traditional strategic understanding of branding is relevant in the digital age. In this chapter, three challenges are discussed: digital consumer journeys, big data, and online brands. These challenges influence branding practices, but do not necessarily invalidate basic branding principles. On the contrary, branding will continue to be important in the future.

KEYWORDS: brand management | digital marketing | social media | big data | consumer journeys

5.1 INTRODUCTION

YouTube star PewDiePie owns the fourth largest channel on YouTube, with 58 million subscribers. In contrast, Coca-Cola, one of the world's oldest and strongest brands, is in 18,465th place, with a mere 300,000 subscribers (Holt, 2016). This example highlights some of the challenges for brands in the digital age. Consumers are increasingly using digital channels for information processing and purchasing decisions; social media and new technologies are creating new digital touchpoints between brands and consumers (Colicev, Malshe, Pauwels, and O'Connor, 2018); and new brand strategies such as influencer marketing and content marketing are being introduced (Olsen and Peretz, 2017). These developments can easily be interpreted as challenging fundamental theories and historical practices in branding. Taylor (2017) calls these developments "the battle of the branding beliefs" and portrays them as a series of dilemmas between the old and the new world. Figure 5.1 illustrates these dilemmas.

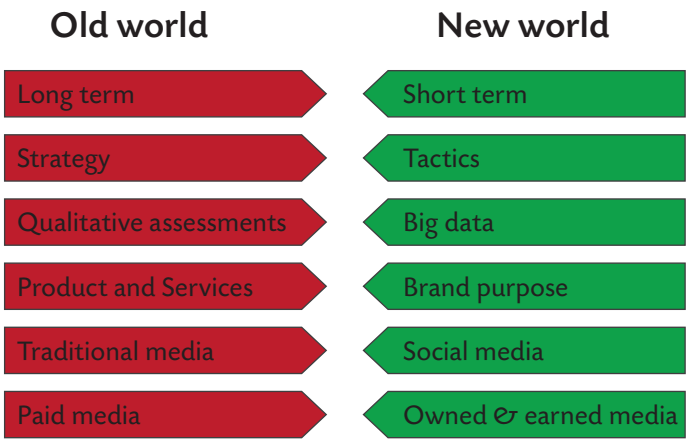

FIGURE 5.1 Battle of the branding beliefs (Taylor, 2017)

According to Taylor (2017), present-day brand management practice tends to be short-term and tactically focused, contrary to the more classic understanding of brands as strategic assets with long-term focus (Keller, 2012). Part of the reason for this change is the availability of behavioral data, such as click patterns. This access to new types of data and new uses of digital media channels fuels a more shortsighted and optimization-focused brand management culture. Mark Ritson at Melbourne Business School comments on this development in Marketing Week: "Marketing seems to be devolving into a base tactical pursuit, devoid of strategic thinking" (Ritson, 2016). Ritson's comment is thought provoking, but also points to an important problem. If the underlying logic of brand management requires long-term strategic thinking, how will short-term digital-focused practices, digital consumer journeys, media channels, and analytic tools affect the branding field?

The purpose of this chapter is to look into these questions and discuss challenges for brands in the digital age. A crucial question is whether fundamental changes in consumer behavior have changed branding practices or whether such changes reflect changes in branding opportunities and communication tools available to brand managers (Batra and Keller, 2016). Many practitioners would claim the former—that new consumer behavior has changed the nature of branding. However, the basic principles of branding have always evolved around two major concepts: identification and differentiation (Samuelsen, Peretz, and Olsen, 2016; Keller, 1993). Identification deals with consumers' abilities to identify the brand in a need situation and to recall the brand from the clutter of other brands and options. Differentiation is the brand's ability to stand out from the crowd and provide relevant and unique benefits to consumers (Keller, Sternthal, and Tybout, 2002). Have

these basic principles really been challenged in the digital age? The answer to that question is not straightforward. On one hand, nothing has really changed. Whether a consumer bases her brand decisions on newspaper ads, Facebook feeds, or Google searches does not change the underlying psychological information processes (Kahneman, 2011), or how consumers complain about bad brand experiences. In the digital age, consumers might be active in online discussion groups and complain about brands on Twitter, whereas their mothers picked up the phone or chatted about bad experiences across garden fences. The underlying human need to gossip, share information (Berger, 2013), and vent negative emotions (Svari and Olsen, 2012) is the same across time and technologies. On the other hand, examples like PewDiePie and Coca-Cola point to an understanding that, even though branding theories may still be valid, they are at least severely challenged in the digital age. In the rest of this chapter, three of these challenges will be scrutinized.

The chapter is organized as follows. First, three major challenges for brands in the digital age will be identified. Each of these challenges will be discussed with a focus on understanding its consequences for branding. Second, the future of branding will be discussed. Will branding continue to be an important strategy in an increasingly digital future? Finally, topics for future research on brand management will be suggested.

5.2 THREE CHALLENGES FOR BRANDS IN THE DIGITAL AGE

Digitalization affects branding practices in many ways. Changes in technologies, market demands, and communication channels provide both opportunities and problems for brand managers. The basic question that is approached in this chapter is whether brands and branding are still relevant in the digital age. This is a rather large question and could be approached in many ways. Essentially, it must be answered by identifying some of the major branding challenges in today's shifting landscape. Three of these challenges are:

1. Digital consumer journeys

Brand practitioners argue that digitalization has changed consumer behavior. Is that right? And if so, what are the consequences for how brands are built, developed, and managed?

2. Big data

New communication channels, technologies, and analytic tools have provided brand managers with enormous amounts of data (big data). How will access to these data affect the balance of short-term optimization of branding activities and the long-term strategic orientation of brand management? Could access to big data actually lead to wrong brand management decisions?

3. Online brands

Consumers are online—for information search, entertainment, status-seeking behavior, and more. Brands must adapt their practices and participate online on consumers' premises. These changes in branding practice can have both advantages and disadvantages for brand management, and can change the way brands are built.

Many other challenges could be discussed, but in this chapter the discussion will be limited to these three challenges. In the next sections they will be discussed in more detail.

5.2.1 DIGITAL CONSUMER JOURNEYS

Consumers use many different communication channels and touch upon brands on many occasions—both online and offline. In many cases, they have the choice of buying brands online. According to eMarketer.com, global ecommerce increased by 5.8% from 2016 to 2017, reaching $22.7337 trillion. In Norway, ecommerce in 2017 was for the first time over 100 billion NOK, an increase of 16% from 2016 (Dibs, 2017). These changes in information search and buying patterns must of course be acknowledged by brand managers. One issue is that digital consumer journeys create an increasing amount of online brand information for consumers, and this information is effortlessly available at all touchpoints with the brand, often with just a swipe of the finger on consumers' smartphones. One could easily conclude that these changes in information availability reduce the importance of brands. However, this information richness is exactly why branding continues to be relevant. Consumers need a sorting mechanism, a way of finding meaning and patterns in the information clutter. Important mechanisms in this process are brand awareness and clear brand positioning (Keller, 1993). Consumers will choose strong brands that are recalled and remembered quickly, and which can provide accessible and relevant benefits in the purchase situation (see Olsen, 2011, for a discussion). This prediction is even truer in a digital age characterized by information overload. For example, Bart, Stephen, and Sarvary (2014) show

how mobile display ads primarily function as reminders of already established brand awareness structures and previously learned brand associations.

This observation closely mirrors classic research by Iyengar and Lepper (2000). In their famous jam study, they found that consumers were ten times more likely to purchase jam on shelf display when the number of jam alternatives was reduced from twenty-four to six. In other words, too many choices turn consumers away.

Another reason why branding remains relevant can be explained by how consumers organize and process information. For example, Daniel Kahneman (2011) has popularized his system 1/system 2 approach. Briefly explained, system 1 operates automatically and quickly, with little or no cognitive effort, and system 2 allocates considerable attention to effortful mental activities. Not surprisingly, strong brands often benefit from a strong system 1 presence. We tend to include strong brands in our daily habits and often automatically evaluate strong brands more favorably than weaker brands. Consequently, strong brands are unconsciously part of our lives. Yet, current digital branding practice, with its focus on rich branded content (Olsen and Peretz, 2017), implicitly assumes that consumers always are in a system 2 mode. However, most brands, products, and service categories are rarely top-of-mind. A critical perspective on digital consumer journeys must therefore ask: Even though consumers in theory have access to all sorts of branded content, how often and when will they use it? Even though digital technology increases access to brand information, consumers do not necessarily appreciate having more information and choices. Thus, brand awareness, brand positioning, and the accessibility of brand associations remain a crucial issue in the digital age.

Summarizing this section, it has been argued that digital consumer journeys have increased the number of consumer-brand touchpoints, changed purchase patterns, and opened up for many brand communication opportunities. Yet, brand awareness, brand association accessibility, and clear brand positioning are still essential. In fact, due to potential information overload, they are even more crucial than ever before.

5.2.2 BIG DATA

In searching for marketing positions in a job database, many titles will be "performance manager", "social media manager", "digital marketing manager", and "content manager." Usually, these positions specify specific skills in Google Analytics and other performance tools. Increased access to big data, and brand managers' shift from traditional to digital media channels, have made these specialized

skills relevant for branding. The possibilities of digitalization are fascinating and at times breathtaking. With these tools it is possible to follow consumers in their digital consumer journeys, measure what they click, see how different tactics (e.g. display ads, search words) convert into sales, what types of digital communication efforts and content are effective, and from which sites consumers enter the brand's webpages and where they go when they leave. The amounts of data are boundless, and clever brand managers can create brand value by optimizing messages and offers. In many ways, the potential to analyze and observe consumer behavior in real time is a revolution for brand management. When we previously experimented with brand messages in analog media channels, it often took weeks before we could measure the results in sales or by other performance indicators. And those indicators were often soft indicators, such as attitudes and beliefs. Today, brand managers can adapt their tactics continuously and instantly measure the effects on facts like sales and customer acquisition.

However, digital analytical opportunities also create new problems for brand managers. Google Analytics and other digital tools focus on the short-term effects of branding efforts. What are the direct consequences of measuring mostly tactical digital behaviors (e.g., click patterns, sales conversion)? In the absence of more strategic brand indicators (e.g., brand associations, brand awareness, brand evaluations, and so on), brand managers run the risk of downplaying the long-term aspect of brand management. Consequently, even with access to more behavioral brand data, they risk losing the strategic perspective of branding and becoming too focused on the short-term. This short-term focus is in itself problematic, since brand management by definition requires a long-term management philosophy (Keller, 1993, 2012). However, it is even more problematic that we risk mistaking consumer data collected via digital tools as equaling consumer insight. It rarely does. We can observe consumer behavior online, but we gain no insight into the reasons as to why consumers behave as they do. Insight demands interpretation and understanding, which take time. Digital performance specialists do not have this time, and often no longer have the skills to follow up on these processes. A challenge for brand managers in the digital age is therefore to balance the possibilities from big data against the time-consuming insight processes of more traditional consumer research. This balance, illustrated in Figure 5.1, will increasingly predict the success of brand management in the future.

To summarize this section, it has been argued that access to big data provides brand managers with a lot of data and opportunities to optimize their branding efforts. However, as managers, we often tend to act upon the data we acquire, which could lead to wrong branding decisions. Access to big data might influence

the delicate balance of short-term optimization and long-term brand management, which in the long run could hurt brands.

5.2.3 ONLINE BRANDS

In December 2017, Stormberg, a producer of outdoor clothing, announced a cooperation agreement with the World Wildlife Fund (WWF). However, as Figure 5.2 shows, the debate soon went sour when consumers engaged in an ongoing public debate about wolves living in Norwegian woods turned against the company. According to the financial newspaper *Dagens Næringsliv* (2 January 2018), consumers opposed to wolves threatened to burn their Stormberg clothes publicly. On Stormberg's Facebook page the debate escalated. Ten days after the announcement, 3,100 comments, 664 shares, 20,000 positive responses, and 142 negative responses were registered. Thus, a majority of Stormberg's customers supported the brand, but the basic lesson in this example is not the valence of the reactions, but the amount, the visibility, and the ease of reporting these reactions.

In a distant era, not so long ago, brand communication was primarily a one-way process. Brand managers and their communication agencies developed creative content, and flooded mass-media channels with advertising messages to influence consumers. Consumers' reactions were both time consuming and technically difficult to express. If consumers vented their reactions to friends and neighbors (e.g., via word of mouth), the number of people they reached was usually quite limited. The rise of social media and digital technology has changed the influence of word of mouth in consumer–brand relationships. Word of mouth has expanded from time-consuming person-to-person processes to an abundance of different online communication formats (Eelen, Özturan, and Verlegh, 2017). These communications include electronic word of mouth (eWOM; Pauwels, Aksehirli, and Lackman, 2016)—e.g., liking a brand on Facebook or writing reviews on electronic forums—posting videos on YouTube and Facebook, retweeting brand messages, and sharing them through social media. Currently, good and bad publicity, mistakes, and brand experiences could potentially be shared and spread to thousands of consumers with only a few clicks. An example is H&M's recent scandal of using a dark-skinned child to advertise a hooded sweater with the text "Coolest Monkey in the Jungle" (Figure 5.2, January 2018). The scandal hit social media and resulted in worldwide outrage toward H&M.

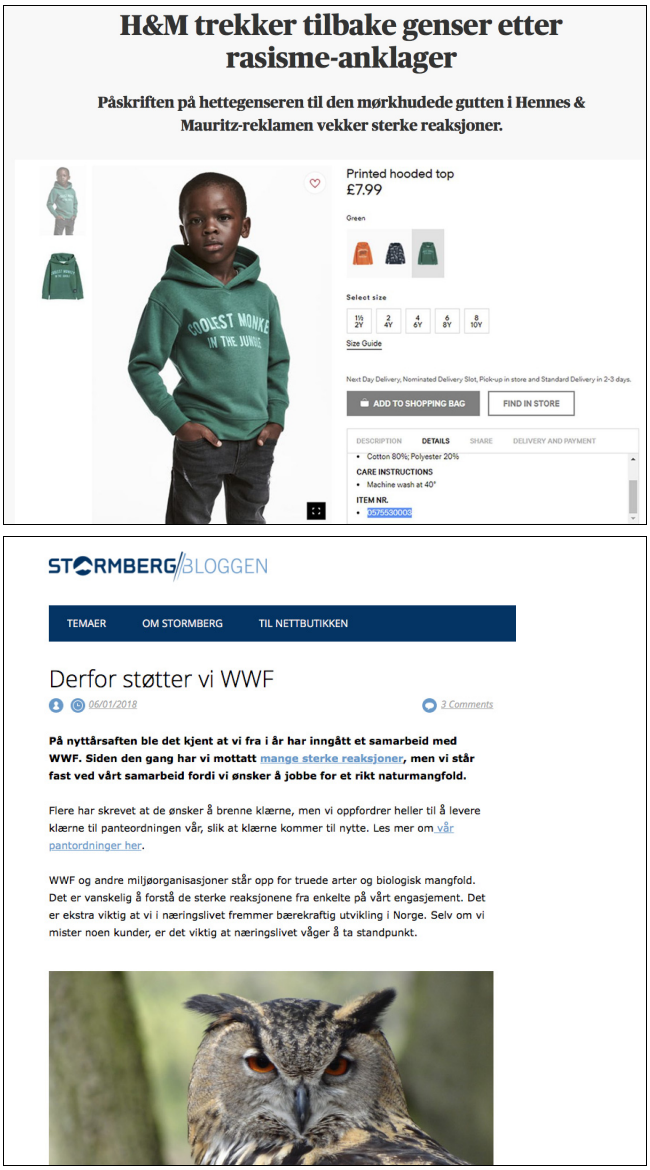

FIGURE 5.2 H&M and Stormberg experience consumer reactions online.

Sources: Aftenposten and Stormberg

Digital technology and media platforms allow for easier consumer–brand relations (Fournier, 2009), but also consumer–consumer relations. By using eWOM, consumers are able to complain, brag, share, and discuss with a minimum of effort. The total effects on brands are therefore formidable, and must be taken seri-

ously by brand managers. Managing the brand–consumer/consumer–consumer dialogues is a critical task for brand managers and can explain why job positions such as "performance managers" and "social media managers" have arisen.

Virality is the buzzword that drives eWOM and brand–consumer interactions online. Akpinar and Berger (2007, p. 318) put it like this:

> Virality has become the holy grail of digital marketing. Rather than focusing on paid media, in which a brand pays to advertise, brands are devoting more and more attention to earned media, in which consumers are the communication channel.

Relevant and interesting topics spread fast, almost like an infectious virus, from consumer to consumer on social media and other digital platforms. Virality helps in creating "brand evangelists" (Kawasaki, 2015), but can also hurt the brand through effective spreading of negative information. These viral effects increase the importance for brands to have a social purpose—to be transparent, sustainable, and conscientious (Olsen and Peretz, 2011). Otherwise the brand will be exposed to negative eWOM in the target group. One example of a brand that has taken this development seriously is Heineken. In May 2017, its "Open Your World" campaign, to promote openness as a value that helps break through the barriers that divide us, spread incredibly fast and has been viewed on YouTube many millions of times. Research from the global consultancy firm BMMG supports this movement and shows that 65% of consumers want to support companies with a strong purpose. However, only 45% of consumers can name such a company (Bemporad, 2017). New digital media provide consumers with tools to find, share, and spread information about companies and brands. Brand managers can no longer just *say* they are responsible for their brand's actions, they must also *act* responsibly and participate in the online dialogue with their consumers.

Another aspect of online brands is *liquid consumption* (Bardhi and Eckhardt, 2017). When was the last time you bought a CD? Fewer and fewer consumers make purchases in this category. Nowadays consumers tend to stream music online, for example by using services like Spotify and Apple Music. New technology opens up for changes in consumption patterns and in many brand categories reduces consumers' needs to own products and brands. Instead of owning the brand, they pay for temporal access and flexibility. Examples are streaming media and video-on-demand services like Storytel and Netflix, which have reduced the need for buying books and DVDs, and new car services like Bilkollektivet, where consumers, for a fee, have access to a range of rental cars. Liquid consumption is

a more flexible way to consume brands, without ownership and using mobile and adapted solutions across situations. Bardhi and Eckhardt (2017) contrast it with the more traditional solid consumption: focus on owning brands with a safe and predicable value across all situations. Figure 5.3 conceptualizes the difference between solid and liquid consumption.

Solid	Liquid
Value resides in size, weight, security, attachement, commitment	Value resides in flexible, adaptable, fluid, mobile, light, detached and fast
Long-standing possessions, attachement/loyalty, identity-related objects Comsumption meaning is stable across contexts	Fluid possessions, lack of loyalty, fewer objects, user value Consumption meaning varies by contexts
Emphasis on ownership and possessions of material objects	Emphasis of access and intangible objects

FIGURE 5.3 The difference between solid and liquid consumption (Bardhi & Eckhardt, 2017)

Liquid consumption has many implications for how we understand brand loyalty and brand relationships (Fournier, 2009). How can you be loyal to a brand you do not possess? So far, examples of brands that have been challenged by liquid consumption have been limited. However, if digitalization continues to alter many markets and business models, it is likely that brands in the future might be significantly affected. It will at least challenge Russell Belk's (1988) classic concept of owning brands as part of consumers' self-identities—the extended self. In general, many established constructs in the branding literature imply brand ownership. If liquid consumption increases, there is a need to reinvestigate many of these constructs.

To summarize this section, it has been argued that the role of brands has changed for online consumers. Currently, it is easier and less time consuming to engage with brands, and the rise of social media and digital platforms changes the nature of brand relationships. The virality of sharing and spreading information has become both an opportunity and a challenge for brand managers. Lastly, brand ownership is declining and consumers are more likely to consume brands in a more liquid manner.

5.3 FUTURE OF BRANDING

New digital technology, media channels, and online consumption patterns influence the art and science of branding. The purpose of this chapter was to discuss how changes fueled by digitalization affect branding practices and brand management. Three challenges to branding in the digital age have been identified: digital consumer journeys, big data, and online brands. The key insights in our discussion have been that these challenges do not reduce the importance of brands, but in many ways shift the practice of branding and provide brand managers with new opportunities. These challenges strengthen the need for deep consumer insight, good understanding of digital consumer journeys, and the importance of well-known and well-positioned brands. To summarize, these three challenges essentially deal with problems of *balance*:

- Balance between opportunities gained from technology and data, and deeply understanding consumer needs
- Balance between short-term optimization and long-term brand strategy
- Balance between short-term sales conversions and long-term consumer brand relationships

Brand managers should certainly embrace new digital opportunities, be curious about new media channels, experiment with new branding activities, and use those that prove to be most effective. However, basic principles of branding and brand management are still, and will continue to be, central. Key concepts like identification, differentiation, and brand positioning (Samuelsen et al., 2016) will not go out of fashion, and may be even more important in the digital future.

REFERENCES

Akpinar, E., and Berger, J. (2017). Valuable Virality. *Journal of Marketing Research*, *54*(2), 318–330.

Bardhi, F., and Eckhardt, G. M. (2017). Liquid Consumption. *Journal of Consumer Research*, *44*(3), 582–597.

Bart, Y., Stephen, A. T., and Sarvary, M. (2014). Which Products Are Best Suited to Mobile Advertising? A Field Study of Mobile Display Advertising Effects on Consumer Attitudes and Intentions. *Journal of Marketing Research*, *51*(3), 270–285.

Batra, R., and Keller, K. L. (2016). Integrating Marketing Communications: New Findings, New Lessons, and New Ideas. *Journal of Marketing*, *80*(6), 122–145.

Belk, R. W. (1988). Possessions and the Extended Self. *Journal of Consumer Research*, *15*(2), 139–168.

Bemporad, R. (2017). *Brand Purpose in Divided Times—Four Strategies for Brand Leadership*. Retrieved from http://bbmg.com/brand-purpose-in-divided-times/.

Berger, J. (2013). *Contagious—Why Things Catch On*. New York, NY: Simon and Schuster.

Colicev, A., Malshe, A., Pauwels, K., and O'Connor, P. (2018). Improving Consumer Mindset Metrics and Shareholder Value Through Social Media: The Different Roles of Owned and Earned Media. *Journal of Marketing*, *82*(January), 37–56.

Dibs (2017). *Norsk E-handel—Alt du trenger å vite om E-handel*. Retrieved from http://info.dibs.no/norsk-ehandel-2017

Eelen, J., Özturan, P., and Verlegh, P. W. J. (2017). The differential impact of brand loyalty on traditional and online word of mouth: The moderating roles of self-brand connection and the desire to help the brand. *International Journal of Research in Marketing*, *34*, 872–891.

Fournier, S. (2009). Lessons Learned about Consumers' Relationships with Their Brands. In D. J. MacInnis, C. W. Park, and J. R. Priester (Eds.), *Handbook of Brand Relationships,* 5–23. Armonk, NY: Society for Consumer Psychology.

Holt, D. (2016). Branding in the Age of Social Media. *Harvard Business Review*, *94*(3), 40–50.

Iyengar, S. S., and Lepper, M. R. (2000), When Choice is Demotivation: Can One Desire Too Much of a Good Thing*? Journal of Personality and Social Psychology*, *79*(6), 995–1006.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kawasaki, G. (2015). Managing Yourself: The Art of Evangelism. *Harvard Business Review*, 93 (5), 108–111.

Keller, K. L. (1993). Conceptualizing, Measuring, and Managing Customer-Based Brand Equity. *Journal of Marketing*, *57*(January), 1–22.

Keller, K. L. (2012). *Strategic Brand Management: Building, Measuring, and Managing Brand Equity 4th Edition*. Upper Saddle River, NJ: Prentice Hall.

Keller, K. L., Sternthal, B., and Tybout, A. (2002). Three Questions You Need to Ask About Your Brand. *Harvard Business Review*, 80(9), 80–86.

Olsen, L. E. (2011). Broad vs. Narrow Brand Strategies—The Effects of Association Accessibility on Brand Performance, *BI Norwegian School of Management—Series of Dissertation 1/2011*.

Olsen, L. E., and Peretz, A. (2017). *Markedskommunikasjon*. Bergen: Fagbokforlaget.

Olsen, L. E., and Peretz, A. (2011). Conscientious brand criteria: A framework and a case example from the clothing industry. *Journal of Brand Management*, *18*(9), 639–649.

Pauwels, K., Aksehirli, Z., and Lackman, A. (2016). Like the ad or the brand? Marketing stimulates different electronic word-of-mouth content to drive online and offline performance. *International Journal of Research in Marketing*, *33*(3), 639–655.

Ritson, M. (2016, May, 11). *Tactics without strategy is dumbing down our discipline.* Retrieved from https://www.marketingweek.com/2016/05/11/mark-ritson-beware-the-tactification-of-marketing/

Samuelsen, B. M., Peretz, A., and Olsen, L. E. (2016). *Merkevareledelse—3 utgave*. Oslo: Cappelen Damm.

Svari, S., and Olsen, L. E. (2012). The role of emotions in customer complaint behaviors. *International Journal of Quality and Service Sciences*, *4*(3), 270–282.

Taylor, D. (2017). *The Brandgym: A Practical Workout for Growing Brands in a Digital Age—3rd Edition*. London, UK: RedDoor Publishing.

DOI: 10.18261/9788215031583-2018-07

Chapter 6 Value Creation, Business Models and Organization Design in a Digital World[1]

ØYSTEIN D. FJELDSTAD AND KNUT HAANÆS

ABSTRACT In most industries, the competitive landscape is rapidly changing and, as a result, companies are speeding through their life cycles at an unprecedented pace. Whole industries are being transformed—media and entertainment, energy—and the changing positions within industries are greater than ever. Digital is the main driver of the current changes. It affects all aspects of how customers behave and how companies create and capture value. We suggest that these technological changes call for the renewed importance in understanding both value creation and principles of organization. Our aim is to address how value creation and organization design is affected by digitization—the opportunities and challenges that digitization presents. We organize our discussion around the concepts "business model" and "organization design" respectively.

KEYWORDS: Business models | value creation | organization design

6.1 BUSINESS MODELS

A business model describes how a company creates and appropriates value (Chesbrough and Rosenbloom, 2002; Teece, 2010). We distinguish between an operational and dynamic dimension of business models (Fjeldstad and Snow, 2018). The operational dimension of a business model describes how a firm creates value for customers by performing activities, and how it appropriates a share of the value created. The operational dimension reflects choices about target customers

1. This chapter draws heavily on prior published works by the authors, which are included in the references. Specifically, Snow, Fjeldstad, Langer, 2017, Fjeldstad and Snow, 2018, and Haanæs and Fjeldstad, 2016.

and product/service offerings, as well as choices about the associated resources and activities. Value created is the difference between customers' willingness-to-pay for the products and services and the opportunity cost of the resources. Value appropriated is the result of bargaining between the participants in larger activity system (Brandenburger and Stuart, 1996; Porter, 1985). Drucker (1954) and Porter (1985) model a business this way.

The dynamic dimension models how the firm changes its business over time. Miles and Snow (1978) describe an "adaptive cycle" in which firms solve the entrepreneurial problem of product/market positioning, the engineering problem of activities and resource configuration, and the administrative problem of balancing exploration and exploitation. Solving the entrepreneurial and engineering problems amounts to developing the operational business model. Miles and Snow identified three purposive adaptation models, which they labeled prospectors, defenders, and analyzers. Prospectors continually develop new products, services, technologies, and markets, i.e. they explore new knowledge (March, 1991). They succeed by moving first relative to their competitors. Defenders leverage their competence in developing process efficiencies for relatively stable product service lines. They search for economies of scale in markets that are predictable and expandable, i.e. they exploit existing capabilities (March, 1991). Analyzers search externally for proven technologies with significant potential for generating new products and services. They use their engineering and operational capabilities to make new products or services better and cheaper, and use their marketing capabilities to commercialize them (Haanæs and Fjeldstad, 2000). Analyzers seek ambidexterity (Tushman and O'Reilly, 2002)—they combine exploration and exploitation, albeit by forgoing radical novelty and optimal efficiency. Finally, the operational model affects the nature of exploration and exploitation because different models imply different capabilities (Fjeldstad and Haanæs, 2001).

Increasingly, firms work with their customers, suppliers, and partners when altering the elements of their business models. Among the key reasons for open innovation business models (Chesbrough, 2006) are access to diverse and situated knowledge (von Hippel, 1994), pooling knowledge resources (Boudreau, Lacetera, and Lakhani, 2011; von Hippel and von Krogh, 2003), and the delivery of products and services that depend on larger platforms and ecosystems (Gawer and Cusumano, 2002). As a result, business models increasingly extend across firm boundaries (Amit and Zott, 2015), and modifying them may affect all or part of the network the firm is embedded within.

6.1.1 A TYPOLOGY OF BUSINESS MODELS WITH IMPLICATIONS FOR DIGITIZATION

There are fundamentally different ways by which firms create value. Each of these also has distinct value propositions, roles of the customer, and mechanisms of value appropriation. That is, they represent different business models.

The value chain (Porter, 1985) is the classical way of modeling the activities of a firm. Its representation fits the typical twentieth century manufacturing process. In contrast, twenty-first century value creation is dominated by firms that create and apply knowledge, and firms that provide global networking services of various types. Such firms are modeled by the Value Shop and the Value Network respectively (Stabell and Fjeldstad, 1998). Furthermore, as a result of extensive efforts to streamline every type of business and focus on the core value creation activities, a fourth type of model has emerged at scale: firms that offer their customers access to shared resources (Fjeldstad and Lunnan, 2015; Haanæs and Fjeldstad, 2016). We briefly review each of these types below, and their particular organizational properties, followed by a discussion about the implications of digitization for each one.

6.1.2 VALUE CHAIN

A value chain transforms inputs into products. The customer is a recipient of the product, which embodies the value created by the firm's transformation process (Ramirez, 1999). Scale, capacity utilization, and the flow of components and products are important to efficient operations, whereas the tailoring of activities to differentiated customer needs is important for value to the customer (Porter, 1985). These counteracting effects on product cost and customer value lead firms to choose between offering a standardized set of products at low cost or targeting differentiated demand with differentiated products. Embodiment of the technology in the product or in the production process is an important value-protection mechanism (Teece, 2010). A value chain forms a sequentially linked value system of suppliers, partners, and customers. Value Chains make trade-offs between cost and differentiation—standardized, one-size-fits-all products can be manufactured at low cost, whereas products targeting differentiated demand can command a premium price but will be more costly to produce (Porter, 1980; 1985). A Value Chain creates and combines components into products, and therefore operations are organized around the flow of components within the supply chain. Value Chains explore new product and process technologies and exploit them by fine-tuning them (Fjeldstad and Haanæs, 2001)

Manufacturing robotics, automation, 3D printing, sensors and digital platforms drive a fourth industrial revolution that allows mass customization—an economy that transcends the traditional trade-off between scale and customization. This revolution changes the forces of globalization by allowing for instant local production and development at low cost. The challenge for industrial firms is twofold. First, they need to embrace digital fast. Second, they need to embrace disruption and not be defensive, as car manufacturers used to be with respect to electric and self-driving cars.

6.1.3 VALUE SHOP

A Value Shop resolves customer problems on a case-by-case basis. Examples are hospitals and consulting firms (Christensen, Grossman, and Hwang, 2008; Christensen, Wang, and van Bever, 2013). Knowledge, and therefore learning, are particularly important to value creation (Løwendahl, Revang, and Fostenløkken, 2001). Problem-solving industries such as medical care typically consist of generalists, who harbor knowledge about a variety of problems, and specialists, who have deep knowledge in a particular area. The client embodies the problem to be solved and may be an active participant in the process of creating solutions (Skjølsvik, Løwendahl, Kvålshaugen, Fosstenløkken, 2007). Value Shops form reciprocally linked value systems of referring, sub-contracting, and collaborating firms that together harness the knowledge required to develop the desired solutions. Status and intellectual property rights, in the form of patents or copyright, safeguard value appropriation. A Value Shop creates and combines competencies to deliver solutions. The organizational design favors mobilization and integration of human and information resources from the network in which the firm is embedded.

Value Shops make trade-offs between providing services that require specialist versus generalist knowledge. This trade-off is reduced with scale (Stabell and Fjeldstad, 1998). Value Shops typically explore by taking on new types of projects and exploit by diffusing knowledge within the firm, reusing it by doing standard work (Fjeldstad and Haanæs, 2001; Løwendahl, Revang, and Fostenløkken, 2001).

Law firms, management consultants, architects, engineering firms, and health care services are all being transformed by digital. Professional services use decision support systems, global databases, collaboration platforms, and communication solutions to improve both the effectiveness and the efficiency of their services. Digital technologies are augmenting and, in some cases, replacing

professionals. They allow one-man firms to be efficient, they allow networks to form instantly, and they allow for the efficient operations of global players like EY, McKinsey, BCG, and Deloitte. Digital is a requirement inside such firms and it is a requirement in their markets. Their clients expect to be working with companies that are one step ahead, not two steps behind.

6.1.4 VALUE NETWORK

A Value Network links nodes—customers, things, and places—and provides services that allow various kinds of exchanges among them. Examples of Value Networks include communication services, transportation (Huemer, LRP,2006), banking and finance, and a wide range of Internet businesses (Afuah and Tucci, 2000). Customers co-produce their own value—but also value for other customers—by making themselves, or nodes that they control, available for networking. Therefore, network scale and composition positively affect the customer value proposition. In many Internet-based network services, there are in addition strong cost economies of scale resulting from low marginal costs associated with each new user or exchange transaction (Varian, 2000). These dual effects of size can create winner-takes-all markets (Shapiro and Varian, 1999). The value systems are vertically layered and horizontally interconnected. Layering allows one service to use another service as its infrastructure. This is common in Internet service ecosystems. Interconnection allows customers of one firm to network with customers of other firms, typical in banking and telecommunications. Lock-in is an important value appropriation mechanism when network externalities affect value creation (Farrell and Klemperer, 2007).

Value Networks create and combine connections among people, places, and things. Operationally, they organize around the platforms that enable those connections and their associated exchanges. They explore new technologies and the relationships that can be serviced by them. They exploit by increasing the size of the networks serviced (Fjeldstad and Haanæs, 2001). Finally, Value Networks trade-off the connectivity and conductivity of their services, i.e. what or who they can connect, and what can be exchanged among them (Evans and Wurster, 2000; Stabell and Fjeldstad, 1998).

Digital technologies enable much more efficient development and operations of networking services. Prominent examples of network service disruptors include Spotify, Netflix, Uber, Airbnb, Amazon, eBay, Facebook, and Google. These companies all outcompeted incumbents by creating highly efficient networking mechanisms that allowed for large scale sharing among users and linking of mul-

tiple product or service suppliers. Their users wanted to be part of the largest network because it provided the best connectivity—the richest offerings, access to the most people and places. These network service firms also all completely outcompeted the incumbents on technology, providing platforms that are able to go to huge global scale fast and drive cost to deliver equally fast.

6.1.5 VALUE ACCESS

A Value Access provider insources the use of physical, informational, and human resources. Classical examples include shared facilities, data processing resources (Brandl, Jensen, and Lind, 2018) or labor pools, for example crewing agencies in shipping (Lorange and Fjeldstad, 2010). Resource access services cover all business functions ranging from IT and HR to facility management and contract manufacturing. By leveraging the total scale in the delivery of a given process across clients, a resource provider can drive up value through reliability, development and quality at the same time as it can drive down costs through shared technology platforms.

Digital is a major enabler because the firm can build global delivery at scale, whilst also offering relevant data and higher delivery quality. It is a scale and platform game, where the winner will be the one who is able to drive scale in customers and utilize technology platforms to deliver. IBM has been there for a long time, as has ISS, Securitas and all ERP players. Whereas the early twentieth century was the time of mass production of products (from cars onwards), the early twenty-first century is the time of mass delivery of services. Digital enables scale, quality, data processing, machine-to-machine communications and sensor-based surveillance.

6.1.6 INTEGRATING MULTIPLE VALUE CREATION LOGICS

A firm may use multiple value creation logics (Stabell and Fjeldstad, 1998; Huemer, 2012). For example, technology development uses a Value Shop logic, whereas distribution uses a Value Network logic. In industries such as pharmaceuticals, software, and entertainment, business models that separate the value configuration logics have emerged. Effective integration and coordination across different value creation logics present important organizational design challenges.

6.2 ORGANIZING DIGITALLY

According to Chandler (1962), "structure follows strategy"—that is, the design of an organization must support its value creation and value appropriation. Both the operational and the dynamic dimension of the business model have implications for organizational design. An organization's coordination and control requirements arise from the value configuration underlying the operational dimension of its business model. Further, exploration favors autonomy within loosely coupled structures, whereas exploitation favors tight coupling between activities and diffusion of best practice (Weick, 1976; March, 1991). Ambidexterity necessitates more complex organizational architectures whereby exploration and exploration are separated temporarily, structurally, or contextually (Junni et al., 2013; O'Reilly and Tushman, 2013; Smith et al., 2010). In addition to affecting business model properties, digital technologies also enable radically new ways of organizing.

Digital technology is not only changing how organizations operate, but also the way we think about organizing. Organizations increasingly include digital and human agents that share the means of communication, control, and coordination. A traditional organization is arranged hierarchically—that is, control and coordination are achieved through an authority (reporting) structure in which superiors plan and coordinate the activities of subordinates, allocate resources, and resolve problems and conflicts (Simon, 1962). A hierarchical organization can be effective in stable and predictable environments because the organization does not have to regularly innovate or adapt to change. Many of today's environments, however, are not stable and predictable; they are volatile, uncertain, complex, and even ambiguous (Johansen, 2007). Such environments are characteristic of knowledge-intensive industries like biotechnology, computers, healthcare, professional services, and national defense.

Digital technology can enable individuals, firms, cities, and governments to become smarter—to expand their capabilities and to adapt to new and changing conditions. As an agile organizational form, the digital organization will be populated with individuals and teams who are accomplished with technology and who can collaborate both inside and outside the organization to make process improvements and develop new solutions.

In previous decades, organizational responses to technological changes were mostly incremental and, in part, enabled by IT improvements that allowed greater scope and dimensionality of organizational control and coordination. Most of those adaptive responses were made within existing hierarchical forms of organizing (Altman, Nagle, and Tushman, 2015). Digital technologies, however, often

disrupt established ways of organizing and require adaptation through collaboration as well as self-organization around shared situation awareness (Endsley, 2000) and knowledge commons (Hess and Ostrom, 2006; Ostrom, 1990, 2010). This adaptive mode is faster and more effective.

Digital technologies play a role in all aspects of operating, controlling, and coordinating the activities of organizations. Broadly speaking, they are used for automating and augmenting tasks, communicating internally among organization members and externally with customers and partners, and in collaborative decision making among digital and human agents (Davenport and Kirby, 2015; Englebart, 1962; Huber, 1990; Licklider, 1960; Simon, 1973).

The technological manifestations of disruption in organization design are clearly visible, as are workplace changes and changes in inter-organizational relationships. What are less visible are changes in the associated design paradigm, which enable organizations to obtain efficiency and effectiveness improvements from investments in digital technology. The new organization design principles are similar to those used in designing digital technologies themselves. Their roots are found in object-oriented systems design (Dahl and Nygaard, 1966) and in the architecture of the Internet (Krol, 1993). In organizational terms, these principles are embodied in actor-oriented architectures, in which the locus of design is actors who collaborate using protocols, processes, and commons (Fjeldstad, Snow, Miles, and Lettl, 2012).

Digital technologies are also used for learning, decision making, and design. Platform companies such as Amazon, Google, Airbnb, and Uber study the data trails of consumer behavior to design markets for greater efficiency, and to build new ones (Lohr, 2016).

Other companies employ digital design tools in collaborating with their customers and partners. Lego provides toolkits on its website that enable entrepreneurs and customers to submit product ideas and start new Lego brick-based businesses (Heinerth, Lettl, and Keinz, 2014).

In traditional organizations, technological artifacts such as manufacturing equipment and computers are controlled by human operators. With the declining costs of global communication and information processing, hierarchy is being by replaced by radically different ways of organizing (Fjeldstad et al., 2012), the digital elements of which include cloud computing, big data analytics, cognitive computing, and collaboration platforms. Artificial intelligence embedded in machinery and tools, as exemplified above, plays an ever-larger role in emerging digital organizations (Kolbjørnsrud, Amico, and Thomas, 2016). The digital organization will need to integrate human and digital agents. The employees and managers will

collaborate with, rather than merely control, the technology in use. Organizing digitally means collaboration with more entities with less reliance on hierarchy for control and coordination. It also entails empowering employees, partners, and customers who use digital tools for the co-creation and co-production of products and services as well as providing digital platforms for self-organized collaboration.

Digital organizations are collaborative, agile, and minimally hierarchical. In many industries, they are populated by human and digital agents who work intelligently side by side. They rely on actor-oriented principles to enable self-organizing, which offers greater connectedness and responsiveness. (Snow, Fjeldstad, and Langer, 2017). Actor-oriented architectures turn complexity into simplicity by using procedures that different types of actors employ to contribute to the overall goal of the organization, identifying the information they need to coordinate their contributions and developing the communication protocols required to interact with one another and with the shared situational awareness.

6.3 CONCLUSIONS

Digital is the main driver of the current changes upending the business world. Digitization affects all aspects of our lives: the way we work, the way we live and the way we consume. We believe that business leaders need to understand clearly how digital enables more effective and efficient business models. Digital transforms organizations. It may radically transform the mechanisms by which activities and resources are differentiated and integrated. For more than two thousand years, we have organized people who use tools and operate machines. The advent of autonomous, intelligent agents with the ability to collaborate also changes our conception of who to organize.

REFERENCES

Afuah, A., Tucci, C. L., 2000. *Internet Business Models and Strategies*: McGraw-Hill Irwin.

Altman EJ, Nagle F, Tushman M (2015) Innovating without information constraints: organization, communities, and innovation when information costs approach zero. In: Hitt MA, Jing Z (eds) *The Oxford Handbook of Creativity, Innovation, and Entrepreneurship*. Oxford University Press, New York, pp 353–379

Amit, R., and Zott, C. (2015). Crafting business architecture: The antecedents of business model design. *Strategic Entrepreneurship Journal, 9*(4), 331–350.

Boudreau, K. J., Lacetera, N., Lakhani, K. R., 2011. Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Science*, 57(5), 843–863.

Brandl, K., Jensen, P. D. Ø., and Lind, M. J. (2018) Advanced Service Offshore Outsourcing: Exploring the Determinants of Capability Development in Emerging Market Firms. *Global Strategy Journal*. In press.

Brandenburger, A. M., and Stuart, H. W. (1996). Value-based business strategy. *Journal of economics and management strategy*, *5*(1), 5–24.

Chesbrough, H., Rosenbloom, R. S., 2002. The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies. *Industrial and Corporate Change*, 11(3), 529–555.

Chesbrough, H. W., 2006. Open innovation: The new imperative for creating and profiting from technology: *Harvard Business Press*.

Christensen, C. M., Grossman, J., Hwang, J., 2008. *The Innovator's Prescription*. New York: McGraw-Hill.

Christensen, C. M., Wang, D., van Bever, D., 2013. Consulting on the Cusp of Disruption. *Harvard Business Review*, 91(10), 106–114.

Davenport TH, Kirby J (2015) Beyond automation. *Harvard Business Review* 94(6):59–65.

Drucker, P.F.,1954. *The Practice of Management*. New York: Harper Brothers.

Endsley MR (2000) Theoretical underpinnings of situation awareness: a critical review. In: Endsley MR, Garland DJ (eds) *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum Associates, Mahwah

Engelbart, DC (1962). Augmenting human intellect: a conceptual framework. Summary Report AFOSR-3223 under Contract AF 49 (638)-1024, SRI Project 3578 for Air Force Office of Scientific Research. Stanford Research Institute, California. Retrieved on March 1, 2007

Evans, P., and Wurster, T. S. (2000) *Blown to bits: How the new economics of information transforms strategy*. Boston, Mass.: Harvard Business School Press.

Farrell, J., and Klemperer, P. (2007). Coordination and lock-in: Competition with switching costs and network effects. *Handbook of industrial organization*, *3*, 1967–2072.

Fjeldstad, Ø. D., and Haanœs, K. (2001). Strategy tradeoffs in the knowledge and network economy. *Business Strategy Review*, 12(1), 1–10.

Fjeldstad, Ø.D. and Lunnan, R. (2015). *Strategi*. Fagbokforlaget.

Fjeldstad, Ø.D. and Snow, C.C. (2018). *Business Models and Organization Design, Long Range Planning*, 51(1) 32–39.

Fjeldstad, Ø. D., Snow, C. C., Miles, R. E., and Lettl, C., 2012. The architecture of collaboration. *Strategic Management Journal*, 33(6), 734–750.

Gawer, A., Cusumano, M.A., 2002. *Platform leadership: How Intel, Microsoft, and Cisco drive industry innovation*. Boston, MA: Harvard Business School Press.

Haanæs and Fjeldstad, 2016 Which business models are most affected by digital? *The Smart Manager*, February, 2017.

Hess C, Ostrom E (eds) (2006) *Understanding Knowledge as a Commons: From Theory to Practice*. The MIT Press, Cambridge.

Hienerth, C., Keinz, P., Lettl, C., 2011. Exploring the nature and implementation process of user-centric business models. *Long Range Planning* 44(5), 344–374.

Huber GP (1990) A theory of the effects of advanced information technologies on organizational design, intelligence, and decision making. *Acad Manag Rev* 15(1):47–71.

Huemer, L. (2012). Unchained from the chain: Supply management from a logistics service provider perspective. *Journal of Business Research*, 65(2), 258–264.

Huemer, L. (2006). Supply management: value creation, coordination and positioning in supply relationships. *Long Range Planning*, *39*(2), 133–153.

Haanæs, K., and Fjeldstad, Ø. (2000). Linking intangible resources and competition. *European Management Journal*, 18(1), 52–62.

Jensen, P. D. Ø. (2012). A passage to India: A dual case study of activities, processes and resources in offshore outsourcing of advanced services. *Journal of World Business*, 47(2), 311–326.

Johansen B (2007) *Get There Early: Sensing the Future to Compete in the Present*. Berrett-Koehler, San Francisco.

Junni, P., Sarala, R. M., Taras, V., and Tarba, S. Y. (2013). Organizational ambidexterity and performance: A meta-analysis. *The Academy of Management Perspectives*, 27(4), 299–312.

Kolbjørnsrud V, Amico R, Thomas RJ (2016). The promise of artificial intelligence: redefining management in the workforce of the future. Accenture Institute for High Performance Business. https://www.accenture.com/_acnmedia/PDF-19/AI_in_Management_Report.pdf. Retrieved on September 19, 2016.

Krol E (1993) FYI on "what is the internet?" Available at: http://www.rfc-editor.org/rfc/rfc1462.txt. Retrieved on November 18, 2016.

Licklider JCR (1960) Man-computer symbiosis. *IRE Transac Hum Factors Electron* 1:4–11.

Lohr S (2016). Goodbye, ivory tower. Hello, candy store. The New York Times Sunday Business (September 4):1–4.

Lorange, P., and Fjeldstad, Ø. D. (2010). Redesigning Organizations for the 21st Century: Lessons from the Global Shipping Industry. *Organizational Dynamics*, 39(2), 184–193.

Løwendahl, B. R., Revang, Ø., and Fosstenløkken, S. M. (2001). Knowledge and value creation in professional service firms: A framework for analysis. *Human relations*, 54(7), 911–931.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization science*, 2(1), 71–87.

Miles, R. E., Snow, C. C., Meyer, A. D., Coleman, J. H. J., 1978. Organizational Strategy, Structure, and Process. *Academy of Management Review*, 3(3), 546–562.

O'Reilly, C. A., and Tushman, M. L. (2013). Organizational ambidexterity: Past, present, and future. *The academy of management perspectives*, 27(4), 324–338.

Ostrom E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, Cambridge.

Ostrom E (2010) Beyond markets and states: polycentric governance of complex economic systems. *Am Econ Rev* 100(3):641–672.

Porter, M.E., 1980. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: The Free Press.

Porter, M. E., 1985. *Competitive Advantage*. New York: The Free Press.

Skjølsvik, T., Løwendahl, B. R., Kvålshaugen, R., and Fosstenløkken, S. M. (2007). Choosing to learn and learning to choose: Strategies for client co-production and knowledge development. *California Management Review*, 49(3), 110–128.

Shapiro, C. V., and Varian, H. (1998). HR (1999) *Information Rules. A Strategic Guide to the Network Economy*. Harvard Business School Press, Boston, Massachusetts.

Simon HA. 1962. The architecture of complexity. *Proceedings of the American Philosophical Society* 106(6). 466–482.

Simon HA (1973) Applying information technology to organization design. *Public Adm Rev* 33(3):268–278.

Smith, W. K., Binns, A., and Tushman, M. L. (2010). Complex business models: Managing strategic paradoxes simultaneously. *Long range planning*, 43(2–3), 448–461.

Snow, C. C., Fjeldstad, Ø. D., and Langer, A. M. (2017). Designing the digital organization. *Journal of Organization Design*, 6(1), 7.

Snow, C. C., Fjeldstad, Ø. D., Lettl, C., and Miles, R. E. (2011). Organizing continuous product development and commercialization: the collaborative community of firms model. *Journal of Product Innovation Management*, 28(1), 3–16.

Stabell, C. B., Fjeldstad, Ø. D., 1998. Configuring value for competitive advantage: on chains, shops, and networks. *Strategic Management Journal*, 19(5), 413–437.

Teece, D. J., 2010. Business Models, Business Strategy and Innovation. *Long Range Planning*, 43(2–3), 172–194.

Tushman, M. L., and O'Reilly, C. A. (2002). *Winning through innovation: A practical guide to leading organizational change and renewal*. Boston, MA: Harvard University Press.

Varian, H. R. (2000). Buying, sharing and renting information goods. *The Journal of Industrial Economics*, 48(4), 473–488.

von Hippel, E., von Krogh, G., 2003. Open source software and the "private-collective" innovation model: issues for organization science. *Organization Science*, 14(2), 209–223.

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative science quarterly*, 21(1), 1–19.

Part II

Contemporary Issues in Governance

DOI: 10.18261/9788215031583-2018-08

Chapter 7
The under-researched family firm: New insights from unique Norwegian data

JANIS BERZINS, ØYVIND BØHREN AND BOGDAN STACESCU

ABSTRACT This chapter reports main findings from a comprehensive study of how Norwegian family firms are governed and how they behave and perform as economic entities. Analyzing all firms from 2000 to 2015, we show that the family firm represents the most widespread way of organizing economic activity, and that family firms differ fundamentally from other firms. Our results suggest that deeper insight into the economics of family firms may make the firms better, and the public debate more informed.

KEYWORDS: Corporate governance | corporate finance | population | family firms | majority control

We have received insightful comments to earlier versions of this chapter from Morten Bergesen, Herman Høyskel, Simone Møkster, Amir Sasson, and Marius Steen.

7.1 MOTIVATION AND OVERVIEW

In this chapter we show that family firms are not like other firms. We also show that the family firm represents the most common way of organizing enterprise in the Norwegian society. Although we are the first to document these properties in the entire population of all firms, there is no reason to think this situation is peculiar to Norway, but rather reflects a global trend.

Nevertheless, the economics of the family firm remains heavily under-researched in any country. This means little is known about how family firms

behave as economic entities. The objective of this chapter is to improve on this unfortunate situation for family firm stakeholders. Owners and employees making important decisions inside the family firm lack systematic insight into typical practice, atypical practice, and best practice. Politicians regulating family firms establish rules with strong impact, but may have only superficial information on what they are regulating. Business school professors training their students can only tell anecdotes about the firms most of the students will be working for or transacting with.

This setting suggests that several stakeholders would benefit from research on the behavior of the family firm in order to understand most firms in the economy better. This chapter contributes to closing this information gap by reporting the main findings from a study of the governance and finance of all Norwegian family firms during the period 2000–2015. These findings come from a project that establishes the broad, overall picture rather than the detailed specifics (Berzins, Bøhren, and Stacescu, 2018b). Accordingly, we will report descriptive statistics rather than tests of behavioral hypotheses, which is the next step in our ongoing project.

In section 7.2, we define the family firm, report why family firms are under-researched, and explain why economists expect family firms to differ from other firms. We describe our database and document the macro-economic significance of family firms in section 7.3. We report governance characteristics in section 7.4, finance characteristics in section 7.5, while we summarize and conclude in section 7.6.

7.2 WHY ARE FAMILY FIRMS UNIQUE AND UNDER-RESEARCHED?

Unfortunately, there is no common definition of the family firm in the literature. More than 90 different definitions exist (European Commission, 2009), making it difficult to compare findings across empirical studies. The lack of a common definition manifests itself as ambiguous conclusions about the main economic relationship analyzed in the family firm literature, which is how performance depends on whether the firm is a family firm (O'Boyle, Pollack, and Rutherford, 2012; Amit and Villalonga, 2014). To illustrate, Villalonga and Amit (2006) find that this relationship between performance and family firm status is either positive, negative, or insignificant, depending on how a family firm is defined. We think an understanding of why some definitions make more sense than others requires an understanding of why family firms are special.

7.2.1 DEFINING THE FAMILY FIRM

We define a family firm as a firm that is majority-owned by individuals related by blood or marriage. This definition reflects both governance and sociology, which are the two dimensions that jointly produce the uniqueness of family firms that we will outline in section 7.2.2.

The governance dimension of the family firm definition reflects that control of the firm's decision making is the fundamental right (Tirole, 2001). Because the owners elect the board, which recruits and replaces the CEO, owners with a majority stake at the shareholder meeting (general assembly) can control every formal governance position without the other owners' consent. Therefore, our definition requires that a group of owners holds more than half the shares. These controlling owners can single-handedly choose its participation intensity in the firm's governance, such as whether to be on the board, be the chair, or the CEO.

The sociological dimension of our family firm definition reflects that we only consider firms where the controlling owners are individuals who constitute a particularly coherent group. We require that the group is tied together by blood or marriage up to the fourth degree of kinship. This means the family also includes great-great-grandparents, great-aunts and great-uncles, aunts and uncles, cousins, grandnieces, and grandnephews.

We prefer this definition of using majority control and sociological coherence to definitions in the literature, which mostly use either control thresholds lower than 50%, sociological thresholds looser than blood or marriage, or governance positions held rather than ownership. For instance, a family firm in Maury (2006) is one where the largest owner has at least 10% of the equity and is either a true family, all the firm's personal owners regardless of the relationship between them, or a private firm. This definition classifies too many firms as family firms from both a control perspective and a sociological perspective.

Other definitions reflect only whether the family holds governance positions, regardless of whether the family is an owner (Anderson and Reeb, 2003; Villalonga and Amit, 2006; Bennedsen et al., 2007). Such definitions ignore ownership altogether, instead using participation in governance as the only threshold.

We think a family firm definition should reflect the family's option to take governance positions, but not whether this option has actually been exercised. Hence, what matters is majority ownership, because it produces the option to govern. This option will presumably be exercised whenever the family finds it optimal to do so. A firm that is majority-owned by a family holding neither a board seat nor the CEO position will not be a family firm if the definition uses only governance positions. Conversely, the firm will be classified as a family firm if the family owns

nothing, but does hold a board seat. In contrast, our definition classifies the first firm as a family firm regardless of the family's participation in governance, but not the latter firm, despite the family's participation. What matters is the right to participate, not actual participation. That right is produced by ownership.

7.2.2 SOURCES OF UNIQUENESS

Family firms are special because the controlling owner is a group of people who are more tightly related sociologically than are other controlling owners. These properties are captured by our definition in section 7.2.1. The important and interesting implication for the economist is that the firm's behavior may reflect the joint maximization of family goals and business goals, which may not be identical. This situation may make characteristics of the owner unusually important for the behavior of the firm. That is, the governance of the firm may depend on the governance of the family controlling the firm (Bennedsen, Perez-Gonzalez, and Wolfenzon, 2010). Three characteristics seem particularly important.

First, demographic characteristics may matter, such as the number, age, and talent of the family members, the family's location, the presence of the founder in the firm's governance, the distribution of ownership within the family, as well as the size, illiquidity, and risk of the family's wealth. For instance, larger families will have a larger pool of family members to select qualified candidates from. Families with illiquid wealth may prioritize dividends before investments. Families with undiversified wealth and income may make the firm behave more conservatively. The family may also make the firm diversify in order to reduce the risk of the overall family portfolio.

Second, family owners may be special because they have information benefits. Family members know each other particularly well after having interacted more or less intensively all their lives. Accordingly, the information asymmetry within the group of controlling owners is unusually small, making it easier to find their best representatives as officers and directors.

The family is also often close to the firm's operations. For instance, the family holds both the chair and CEO positions in 79% of Norwegian family firms (Bøhren et al., 2018). This means the owner is unusually well informed about the firm's future performance. The resulting low information asymmetry between the firm and the controlling owner reduces the family firm's cost of capital (Leland and Pyle, 1977; Anderson, Mansi, and Reeb, 2003).

Third, family firms may be special due to private benefits, which accrue to the controlling family only (Jensen and Meckling, 1976). Examples of private bene-

fits are when a firm with the family's name has high reputation in society (social prestige), when a family-controlled newspaper influences common opinion (political impact), when the firm employs family members with lower skills than outside candidates have (nepotism), and when family members trade with the firm at below-market prices (tunneling).[1]

Just like demographics and information advantages, private benefits may influence the firm's behavior. A feeling of pride for the family firm's name, and loyalty to the founders, may make the firm's survival particularly important to the controlling family. This concern for survival may make the family firm more long-termist and more patient than other firms in its investment, financing, and employment decisions (Sraer and Thesmar, 2007). Concerns for survival may also make the family firm adopt less aggressive growth strategies and choose industries and products with less risk (Almeida and Wolfenson, 2006).

7.2.3 REASONS FOR LOW ATTENTION

The existing research on corporate governance (Hermalin and Weisbach, 2017) and corporate finance (Eckbo, 2007) is heavily biased toward firms that are public (listed on a stock exchange and widely held) rather than private (not listed and closely held). Because almost every family firm is private, the lack of research on private firms carries over to family firms.

There are at least three reasons why economists have paid limited attention to private firms in general and to family firms in particular. First, public firms may look more attractive because the quality of the firm's behavior may be measured by the observable market value and not just by the book (accounting) value, which is normally the only option in private firms. Thus, performance is harder to measure when the firm is private.

Second, regulation puts stronger requirements on the external communication of public firms than of private firms. Public firms must regularly publish standardized, audited accounting statements that data vendors make easily accessible to investors, analysts, and researchers worldwide. Reliable accounting data for private firms are harder to obtain in most countries. Correspondingly, governance data for public firms are easily accessible, but not for private firms. Moreover, even if governance data had been available, family firms cannot be identified and analyzed reliably

1. These examples illustrate that private benefits may or may not be costly for the minority owners. The first two examples may not produce negative consequences for them, while the third and fourth do. Thus, private benefits increase the family's utility of controlling the firm, while the effect on the other owners is neutral or negative.

without knowing the family relationships between owners, directors, and CEOs. This task requires census data for larger samples, which are not publicly available.

Finally, economists tend to consider the family firm an anachronism (Bennedsen, Perez-Gonzalez, and Wolfenzon, 2010). The reason may be a lack of recognition of the fact that family firms continue to play a strong economic role around the world (Franks, Mayer, and Rossi, 2009; Mehrotra et al., 2013). This prevalence of the family firm, despite lower frictions in capital and labor markets, may jointly refute the idea that the family firm is a viable organizational form only in underdeveloped markets (Khanna and Yafeh, 2007). Moreover, most governance researchers may simply have overlooked the family firm because the ruling paradigm concerns the widely held firm, and the resulting separation between weak owners and strong managers (Berle and Means, 1932; Roe, 1994; Hermalin and Weisbach, 2017).

Unfortunately, existing findings from public firms may not apply to family firms. We will establish that family firms face different environments and have different governance. For instance, family firms are almost always private and cannot finance themselves in active equity markets. This means their shares can be traded only at high transaction costs, and that their minority shareholders enjoy less regulatory protection than in public firms. Our data show that family firms have unusually concentrated ownership, small boards, and many owners in CEO and chair positions. Theory suggests that such characteristics matter for the firm's investments, capital structure, dividend policy, growth, and risk management. This behavior may in turn influence performance, such as the return on capital invested. The literature has just started addressing the relationship between performance and the sources of family firm uniqueness. The only exception is the small subsample of public family firms, which may differ strongly from their private counterparts.

7.3 DATA, GROUPS, AND MACRO-ECONOMIC SIGNIFICANCE

Our database captures the entire population of Norwegian firms with limited liability (*AS* and *ASA*) during the period 2000–2015. Many private firms are organized in corporate groups with parents and subsidiaries. We report a firm separately if it has no majority parent, reporting one observation for a set of firms organized as a corporate group by consolidating the firms' activities.[2]

Experian (www.experian.no) provides the accounting, ownership, and board data, the census data on family relationships are from the National Registry (folke-

2. We find that groups controlled by families are up to eight levels deep and may contain almost 100 subsidiaries.

registeret) (www.skatteetaten.no/en/person/national-registry), and the Centre for Corporate Governance Research (www.bi.edu/ccgr) organizes these two data sets as one integrated database. Tax return data on shareholder wealth and income are from Statistics Norway (www.ssb.no/en) and the Norwegian Tax Administration (Skattedirektoratet) (www.skatteetaten.no). The average sample contains about 86,000 family firms and nonfamily firms per year. The detailed construction of the sample is described in chapter 4 in Berzins, Bøhren, and Stacescu (2018b).

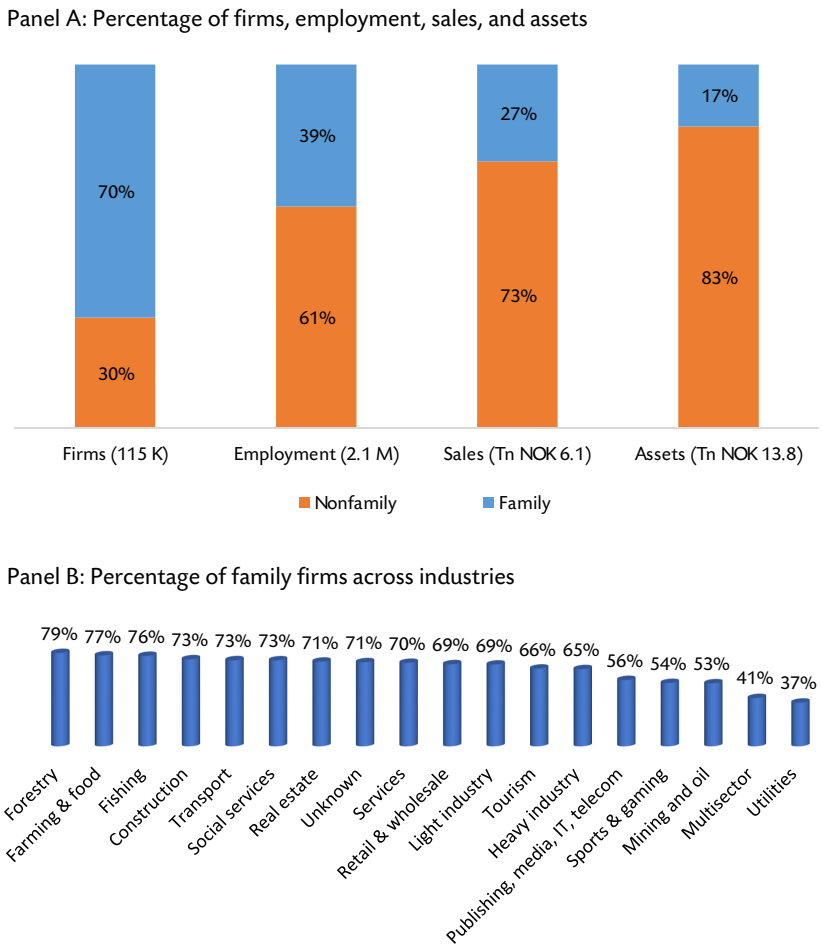

FIGURE 7.1. The macro-economic significance of family firms in 2015
This figure shows the prevalence of family firms in the Norwegian economy. A family firm is majority-owned by individuals related by blood or marriage up to the fourth degree of kinship. Nonfamily firms are all other firms. The sample is all Norwegian firms with limited liability by year-end 2015.

Figure 7.1 shows how family firms contribute to economic activity in the Norwegian economy in 2015. According to panel A, 70% of all firms are family firms, and they account for 39% of aggregate employment, 27% of aggregate sales, and 17% of all assets. Panel B shows that family firms are more common in certain industries. The fraction of family firms varies between 79% and 37%. For instance, 69% of the firms are family-controlled in retail and wholesale, which also has the largest number of family firms in the economy (22,780 firms; not shown in Figure 7.1). Only 53% are family firms in mining and oil, which also has few family firms (432 firms; not shown in Figure 7.1).

These numbers document that the family firm is by far the most common way of organizing an economic enterprise. Because the share of firms exceeds the share of employment, sales, and assets, family firms tend to be smaller and more labor intensive than nonfamily firms. Family firms also seem to gravitate toward industries that use more labor and less assets. These findings may suggest that, compared to nonfamily owners, family owners are more constrained by limited funds, more reluctant to grow even if they could, more often choose industries with lower capital requirements, and more often specialize in managing labor. We address these questions using corporate finance data in section 7.5.

7.4 CORPORATE GOVERNANCE

Corporate governance concerns the relationship between owners, directors, and officers (Shleifer and Vishny, 1997). We count the family as one owner by adding each family member's equity stake in the firm into the family's stake. To capture differences across subgroups of family firms, we show the results separately for all family firms, for sole entrepreneurships versus classic firms, and for large versus medium and small firms.

We define sole entrepreneurships as firms less than ten years old and controlled by one person.[3] The idea is to isolate the case where one individual in the family rather than several start the firm. Sole entrepreneurships may have less family firm flavor than other family firms, which we call classic. We define a large firm as having at least 10 employees and sales above NOK 10 million at 2015 purchasing power on average over the sample period. Groups are reported as one observation, and we carefully measure governance and finance characteristics at the appropriate level in the group.

3. According to this definition, only one family member can own shares in an entrepreneurship.

We analyze sample characteristics of the ownership structure in section 7.4.1, the board in section 7.4.2, and the CEO in section 7.4.3. While we consider all family firms as one sample in sections 7.4.1–7.4.3, we compare subsamples of family firms in section 7.4.4.

7.4.1 THE OWNERSHIP STRUCTURE

Our database includes all firms in the Norwegian economy. Therefore, we can measure ultimate ownership, which is the owner's direct equity stake plus the indirect stake owned through corporate intermediaries. Accounting for indirect ownership is particularly important in our sample period, as the number of holding companies more than tripled in 2005, when increased dividend taxation for persons discouraged the use of direct ownership (Berzins, Bøhren, and Stacescu, 2018a). Similarly, ship-owners reorganized their corporate groups in order to adapt to the new tax regime for ship-owning firms in 2007.

Panel A of Table 7.1 shows properties of the ownership structure in 2015. The most common concentration measure is the largest equity stake, which exceeds 50% in all family firms by definition. Counting the family as one owner, the average largest owner holds 50% in nonfamily firms and 95% in family firms. Thus, while the average largest owner in both firm types has simple majority (1/2), the largest owner in family firms has supermajority (2/3) by a wide margin. To illustrate, while the largest average owner can single-handedly elect the board in both firm types, that owner can also amend the charter in family firms. In fact, the controlling family can even buy out the minority owners, which requires a 90% stake (Bøhren and Krosvik, 2013).

TABLE 7.1 Corporate governance characteristics of Norwegian firms in 2015

			Family firms				
	All firms	Non-family firms	All	Classic	Sole entre-preneurships	Large	Medium & small
Panel A: Ownership							
Largest owner	83%	50%	95%	94%	96%	90%	95%
Number of owners	3.05	5.74	1.87	2.42	1.28	3.22	1.69
Second largest owner	15%	24%	11%	18%	3%	16%	10%
Inside owners	88%	67%	94%	91%	97%	85%	95%
Percent single-owner firms	63%	14%	81%	75%	88%	65%	83%
Panel B: Board							
Family is on the board			99%	99%	98%	98%	99%
Family's share of seats			91%	91%	90%	86%	91%
Family has board majority			85%	87%	84%	77%	87%
Family has all seats			81%	80%	83%	67%	83%
Family has chair			80%	89%	70%	80%	81%
Director ownership	79%	66%	83%	88%	76%	82%	83%
Chair ownership	58%	26%	66%	63%	68%	58%	67%
Minority owner has board seat			8%	9%	6%	11%	7%
Board size	2.08	2.98	1.64	1.84	1.37	2.11	1.57
Director turnover	3%	7%	2%	2%	2%	3%	2%
Proportion female directors	21%	19%	22%	22%	22%	17%	22%
Proportion female directors in 2000	16%	12%	17%	19%	15%	13%	18%
Age female directors	49.9	49.2	50.3	51.7	46.1	51.6	50.1
Age male directors	52.6	51.6	53.1	55.3	48.3	54.5	52.9
Standard deviation director age	8.2	7.2	9.3	9.8	7.8	10.1	9.0
Standard deviation director gender	33%	29%	35%	33%	37%	29%	35%
Panel C: CEO							
Family has CEO			83%	82%	85%	73%	85%
Family has CEO and chair			65%	74%	61%	58%	67%
CEO ownership	73%	37%	83%	72%	96%	68%	85%
CEO has majority	64%	0%	81%	66%	98%	64%	82%
Proportion female CEOs	19%	18%	19%	18%	20%	10%	21%
Proportion female CEOs in 2000	14%	12%	14%	14%	14%	7%	15%
Age female CEOs	46.9	45.7	47.3	49.9	44.1	50.1	47.2
Age male CEOs	50.0	48.6	50.6	53.9	45.9	52.1	50.4
Number of firms	115,259	35,077	80,182	32,281	32,422	9,560	70,622

This table shows ownership, board, and CEO characteristics of all Norwegian firms with limited liability in 2015. All figures are mean values. A family firm is majority-owned by individuals related by blood or marriage up to the fourth degree of kinship, while nonfamily firms are all other firms. A sole entrepreneurship is less than ten years old and is controlled by one person, while classic firms are the remaining family firms. A large firm has average sales above NOK 10 million and at least 10 employees, while the remaining family firms are medium and small. Single-owner firms have just one owner, which may be a family with several individual owners. Ownership is measured as the sum of the shareholder's direct and indirect equity holdings in the firm. "Inside owners" is the equity fraction held by the firm's officers and directors as a group. Business groups represent one observation each.

Although both firm types in our sample have few owners on average, there are still considerably fewer owners in family firms than in nonfamily firms (1.87 vs. 5.74, respectively). The average stake of the second largest owner is smaller, being 11% in family firms and 24% in nonfamily firms. Moreover, while 81% of the family firms have no other owners than the family, such single-owner firms constitute only 14% of the nonfamily firms. Finally, officers and directors, who are what we call the firm's insiders, own 67% in nonfamily firms and 94% in family firms.

The figures in panel A reflect two properties of the family firm's ownership structure with important implications for governance (Edmans and Holderness, 2017): Ownership concentration is very high, and insider ownership is also very high. These two properties speak directly to what the literature calls the first and the second agency problem, respectively (Villalonga and Amit, 2006). The first agency problem concerns conflicts of interest between the firm's owners and insiders (Shleifer and Vishny, 1997). This alignment problem is less serious when insiders have incentives to act as the owners would have done themselves (Demsetz and Lehn, 1985). That automatically happens when the insiders are in fact large owners, as in our case. Therefore, the first agency problem is mostly negligible.

The second agency problem is potentially serious when ownership is concentrated (La Porta et al., 2000). This situation may tempt the controlling family to make decisions that benefit themselves at the other owners' expense. However, this problem is smaller the more the controlling stake exceeds the minimum level of 50%. For instance, a family owning 51% pays only 51% of the loss they cause the firm that underprices goods to the family. The remaining 49% must be paid by the minority. In contrast, a family owning 99% pays 99% of the loss, while the minority pays only 1%. Panel A shows that the average family is rather close to the latter scenario of holding a very high control stake. Moreover, there is no second agency problem whatsoever in 82% of the family firms, because they have no minority owners.

We conclude that, compared to other firms, the ownership structure of family firms makes them less exposed to conflicts of interest both between owners, officers, and directors, and also less exposed to conflicts between large and small owners.

7.4.2 THE BOARD

Panel B of Table 7.1 shows board characteristics.[4] The controlling family is on the board in 99% of the family firms, holds at least half the seats in 85%, every seat in 81%, and the chair in 80%. The directors own more equity in family firms than in nonfamily firms (83% vs. 66%). The tendency is even stronger for the chair (66% vs. 26%).

This situation means that, just like the shareholder meeting in panel A, the board meeting is totally dominated by the controlling family. In fact, the family's average fractions of share capital and of board seats are quite close, being 95% and 91%, respectively. Thus, once more, we find that the first agency problem is minuscule in family firms. However, the composition of the board does not mitigate the potential second agency problem. Although minority owners hold a higher fraction of seats than of shares (9% vs. 5%), their formal power is mostly nil because the board always decides by simple majority.

Family firms tend to have smaller boards than nonfamily firms (1.64 vs. 2.98 seats) and directors with longer tenure (2% vs. 7% likelihood of at least one director being replaced in a given year).[5] Female directors are more common in family firms (22% vs. 19% of the seats), and more common than fifteen years earlier in both firm types. The average female director is about two and a half years younger than males in both firm types. Finally, diversity as measured by the standard deviation of director age and of gender both suggest that family firms have the more heterogeneous board.

Taken together, these figures reflect that the family is very much present on the family firm's board. This active involvement reduces the separation between ownership and control, which is positive from a corporate governance perspective. The potential problem is conflicts with minority owners, who seldom have formal power in the boardroom.

7.4.3 THE CEO

Panel C shows that the family's dominance in shareholder meetings and boardrooms carries over to the CEO position. For instance, the family holds the position of CEO in 83% of the firms, and both the chair and CEO positions in 65%. Again, the first agency problem is practically nonexistent, while there is no obvious mitigation of the second.

---

4. Because our data cannot tell how nonfamily directors are related to the owners, the table has no entries for seat identity in nonfamily firms.
5. Norwegian boards are very small by international standards. For instance, the average size of 6.2 including employee directors in public firms is among the smallest boards in the world (Bøhren and Strøm, 2010).

The proportion of female CEOs is close to 20% in both firm types, increasing by about five percentage points over the sample period. Both male and female CEOs are on average about two years older in family firms than in nonfamily firms.

7.4.4 TYPES OF FAMILY FIRMS

We next compare governance characteristics across different types of family firms. The figures in the four rightmost columns of Table 7.1 show that the sole entrepreneurship and the large family firm differ more from the average family firm than do classic or middle-sized and small family firms. Because sole entrepreneurships and large family firms are also the ones that differ the most from each other, we compare these two family firm types in the following.

Regarding ownership, the ownership concentration and the insider holdings are both highest in sole entrepreneurships, and lowest in large firms. As for the board, family domination is strongest—and size smallest—in sole entrepreneurship, while the opposite is true for large firms. Both types use external chairs more often than other family firms do. The CEO in sole entrepreneurship is more often a controlling owner, while large firms are at the opposite end. Sole entrepreneurships use female CEOs more often than do large firms. They also have the youngest directors and CEOs regardless of gender, while large firms have the oldest.

Summarizing section 7.4, we find that the conflict potential between owners, directors, and officers is very small in family firms due to their ownership structure, board composition, and the CEO's background.[6] Conflicts between the family and minority owners are also mitigated by the ownership structure, but not by the background of officers and directors. This conclusion holds across different types of family firms, although sole entrepreneurships and large firms have governance that resembles the widely held public firm the least and the most, respectively. The very strong family dominance in board and CEO positions may create settings where the beneficial effects of family control are offset by the negative effect of recruiting officers and directors from a pool of talent that is too limited.

7.5 CORPORATE FINANCE

In this section, we briefly describe micro-economic characteristics of the family firm in 2015 as summarized in Table 7.2. Like in section 7.4, we first compare all

6. This situation differs widely from the widely held public firm, where large owners are mostly much smaller, owners are much less present on the board, and the CEO is very rarely a large owner.

family firms as a group to nonfamily firms in sections 7.5.1–7.5.3, comparing different types of family firms to each other in section 7.5.4.

TABLE 7.2 Corporate finance characteristics of Norwegian firms in 2015

			Family firms				
	All firms	Non-family firms	All	Sole entrepreneurships	Classic	Large	Medium & small
Panel A: Size							
Means							
Sales (million NOK)	53.1	128.0	20.3	9.2	29.6	132.5	5.1
Assets (million NOK)	124.1	339.6	29.8	5.3	38.1	205.0	6.1
Employees	19.3	38.2	11.1	6.0	15.4	53.7	5.3
Medians							
Sales (million NOK)	3.3	5.6	2.7	1.8	4.0	32.6	2.2
Assets (million NOK)	2.2	3.4	1.9	0.9	3.4	23.5	1.4
Employees	4.0	6.0	3.0	2.0	4.0	21.0	2.0
Panel B: Factor intensity							
Means							
Assets per employee (million NOK)	4.53	7.84	3.08	1.38	4.17	3.94	2.97
Sales per employee (million NOK)	1.98	2.89	1.58	1.26	1.83	2.58	1.45
Sales to assets	2.78	3.24	2.58	3.20	2.09	1.56	2.72
Panel C: Financing							
Means							
Debt to assets	0.75	0.79	0.73	0.80	0.67	0.64	0.74
Short- to long-term debt	4.32	4.94	4.04	4.01	3.93	4.06	4.03
Cash to assets	0.31	0.30	0.32	0.36	0.27	0.15	0.34
Payout ratio	18.8%	19.7%	18.4%	14.9%	20.8%	23.2%	17.6%
Payout ratio for payers	80.0%	82.4%	78.9%	79.9%	77.7%	64.3%	83.0%
Proportion of payers	16.8%	16.0%	17.1%	13.0%	20.3%	31.0%	15.2%
Panel D: Growth, risk, and profitability							
Means							
Sales	6.0%	6.5%	5.7%	10.4%	3.5%	8.6%	5.3%
Assets	4.9%	3.6%	5.4%	10.2%	3.1%	9.7%	4.7%
Risk	0.26	0.25	0.27	0.27	0.26	0.21	0.28
Return on assets (ROA)	3.6%	0.6%	5.0%	4.3%	5.7%	9.4%	4.2%
Number of firms	115,259	35,077	80,182	32,281	32,422	9,560	70,622

This table shows behavioral characteristics of all Norwegian firms with limited liability in 2015. A family firm is majority-owned by individuals related by blood or marriage up to the fourth degree of kinship, while nonfamily firms are all other firms. A sole entrepreneurship is less than ten years old and is controlled by one person, while classic firms are the remaining family firms. A large firm has average sales above NOK 10 million and at least 10 employees, while the remaining family firms are medium and small. "Payout ratio" is dividends divided by earnings, while "Risk" is the standard deviation of sales divided by average sales in 2013–2015. Every ratio in panel C except "Fraction payers" is winsorized at the top 2.5% of the distribution. In panel D, the first two ratios are winsorized at the top 5%, while ROA is winsorized at the bottom and top 2.5% of the distribution. Business groups represent one observation each.

7.5.1 FIRM SIZE AND FACTOR INTENSITY

We measure size in panel A by sales, assets, and employees, reporting both means and medians. The numbers show at the firm level what panel A of Figure 7.1 shows in the aggregate, i.e., that family firms tend to be smaller than nonfamily firms. The average family firm is about one-tenth the average nonfamily firm according to sales or assets, and about one third according to employment.

However, mean values for nonfamily firms in particular are heavily influenced by some very large firms. This effect is evident when we compare the medians, which is the observation at the center of the distribution (half the firms are above the median and half the firms are below): All three size measures suggest that the typical firm is much smaller than what the means suggest. Moreover, the typical family firm is about half the size of the typical nonfamily firm. Specifically, the median family firm employs three people and sells for NOK 2.7 million, while the median nonfamily firm employs six people and sells for NOK 5.6 million The corresponding figures for assets are NOK 1.9 million and NOK 3.4 million, respectively.

The skewness toward low size can be further illustrated by considering the shape of the distribution in more detail. For instance, half the family firms have sales between NOK 1 million and NOK 8 million, while one-tenth has sales above NOK 28 million.

The factor intensities in panel B confirm the impression that family firms are less capital intensive than nonfamily firms are. For instance, the average family firm has assets of NOK 3.08 million per employee, while the average nonfamily firm has NOK 7.84 million.

7.5.2 FINANCING

Panel C shows that the family firm is on average somewhat less debt-financed than nonfamily firms, as the debt to asset ratios are 0.73 and 0.79, respectively. Moreover, family firms tend to have less debt with short maturity, the ratio of

short-term to long-term debt being 4.04 and 4.94, respectively. As the average ratio of cash to assets is practically equal in the two firm types (about 0.30), family firms tend to have the higher working capital.

The standard measure of payout from the firm to its owners is the payout ratio, which we calculate as dividends divided by earnings. The figures suggest that the dividend policy is very similar in the two firm types. About 17% of the firms pay dividends, about 19% of the earnings are paid out if we also include the firms that do not pay, while 80% of the earnings are paid by the subsample of firms that do pay.[7]

7.5.3 GROWTH, RISK, AND PROFITABILITY

The average annual growth rate in sales and assets is shown in panel D. Family firms grow less than nonfamily firms do as measured by sales (5.7% vs. 6.5%), while they grow more as measured by assets (5.4% vs. 3.6%).

We measure risk by the coefficient of variation for sales, which is the standard deviation of sales divided by average sales. Using the past three years as the measurement period, we find that family firms and nonfamily firms have very similar risk on average, the coefficient of variation being 0.27 in family firms and 0.25 in nonfamily firms, respectively.

Finally, we measure profitability by return on assets (ROA), which is operating earnings after tax but before interest plus interest divided by total assets. Family firms outperform nonfamily firms by a large margin, the average ROA being 5.0% and 0.6%, respectively. If we compare the medians rather than the means, and use the entire sample period 2000–2015 rather than just 2015, the ROA is 1.8 percentage points higher in family firms than in nonfamily firms.

This ROA difference, which may be called the family firm premium, resembles the level reported for public family firms internationally (Amit and Villalonga, 2014). A comprehensive analysis is required to understand this premium. This analysis will start from the fact that the premium in Table 7.2 ignores any firm characteristic other than firm type (family vs. nonfamily) The premium also ignores any variation in ROA across firms within a firm type, using only the average. There may be at least four explanations of the premium that are not mutually exclusive.

7. The dividend tax increase in 2005 had a very strong effect on dividend policy in both family firms and nonfamily firms. The average payout ratio dropped from about 80% to about 15%, and the fraction of dividend payers dropped from about 50% to 15%. These numbers have increased somewhat during the last five years.

First, family firms and nonfamily firms may have different accounting practices. For instance, family firms may be more conservative in valuing their physical assets, and may less often account for intangible assets. Second, because Figure 7.1 shows that family firms are unevenly distributed across industries, the premium may be due to differences in industry returns. Third, the premium may be due to corporate finance characteristics discussed in this section, such as the firm's size, growth, financing, and risk. Fourth, the explanation may be corporate governance and the unique properties of family ownership we discussed in section 7.4 and quantified in Table 7.1. Our ongoing project addresses these various explanations in order to understand better why family firms behave and perform differently from nonfamily firms (Berzins, Bøhren, and Stacescu, 2018).

7.5.4 DIFFERENT FAMILY FIRMS

Like Table 7.1 on corporate governance, Table 7.2 on corporate finance shows that the atypical firms among the family firms are the sole entrepreneurships and the large firms. Sole entrepreneurships have unusually high labor intensity, high debt, and low dividends, while large firms are at the opposite end. Large firms take less risk, while both firm types have the highest growth and profitability among the family firms.

Summarizing section 7.5, the frequency distribution of size for the population of all firms is heavily skewed toward low size. The median family firm is half the size of the median nonfamily firm, and family firms are more labor intensive than nonfamily firms are. The average family firm has lower financial leverage and higher working capital, while the dividend policy is very similar in the two firm types. Family firms have higher average asset growth, but lower sales growth. Average risk is practically independent of firm type, while the average family firm outperforms the average nonfamily firm. As for corporate governance, the sole entrepreneurship and the large firm are the outliers among the family firms in corporate finance as well. One common property is high average growth and profitability.

7.6 SUMMARY AND CONCLUSION

Surprisingly little is known internationally about the family firm as an economic entity. This situation is worrying when considering a main finding we report in this chapter. Using novel data over sixteen years from the population of Norwegian firms and their owners, we find that the family firm is the dominating organiza-

tional form in the economy, and that family firms behave and perform differently from other firms. Family firms represent 70% of all firms, employ four-tenths the labor, generate one-quarter of the sales, and own about one-sixth of the assets.

We define a family firm as being majority-owned by individuals related by blood or marriage. Family firms are special because the controlling owner is a group of people who, sociologically, are unusually tightly related. This means the governance of the family firm may depend on the governance of the family as reflected in its demographics, information advantages, and private benefits. The family's incentives to ensure firm survival may make the family firm more long-termist, risk averse, and patient in its decision making than other firms.

We find that the potential conflict of interest between owners, directors, and officers is very small in most family firms due to their ownership structure, board composition, and CEO background. Conflicts between family and nonfamily owners are also mitigated by the ownership structure, but not by the background of officers and directors. This is true across different types of family firms, although we find that sole entrepreneurships and large family firms have the governance that resembles the widely held public firm the least and the most, respectively. The very strong family dominance in shareholder meetings, boardrooms, and CEO positions may create settings where the beneficial effects of family involvement are offset by the negative effect of recruiting governance resources from too limited a talent pool.

The typical family firm is half the size of the typical nonfamily firm, and both firm types consist of much more small firms than large. Family firms are more labor intensive, have lower leverage and higher working capital, while the dividend policy is very similar. Family firms have higher asset growth, but lower sales growth, while the risk is practically identical. Family firms outperform nonfamily firms on average as measured by returns to assets. The sole entrepreneurship and the large firm are atypical among the family firms regarding both behavior and performance.

These findings reflect a first, rough attempt at answering important questions using new and complicated data. As far as we are aware, no existing study has used population data to explore such a wide range of economic characteristics for family firms. This setting implies that our findings should be used with caution. We think a fruitful way forward is to use our results to ask deeper questions and design more careful analyses. A top priority is to explore where the performance premium for family firms comes from. One explanation is that the premium is simply due to different accounting practices in family firms than in nonfamily firms. A second explanation is our finding that family firms are overrepresented

in certain industries. The premium may also be due to the fact that we have only considered averages and ignored the differences in behavior and performance across individual firms. Finally, the premium may be due to our finding that the family firm's governance has unique properties rooted in the ownership structure and the controlling family. Our ongoing project addresses these explanations in order to understand better why family firms are different.

REFERENCES

Almeida, H., and D. Wolfenzon, 2006. Should business groups be dismantled? The equilibrium costs of efficient internal capital markets. *Journal of Financial Economics* 79, 99–144.

Amit, R., and B. Villalonga, 2014. Financial performance of family firms. In L. Melin, M. Nordqvist and P. Sharma (Eds.), *The SAGE Handbook of Family Business*, 157–178. London: Sage.

Anderson, R., and D.M. Reeb, 2003. Founding family S&P 500. *Journal of Finance* 58, 1301–1329.

Anderson, R., S.A. Mansi, and D.M. Reeb, 2003. Founding-family ownership and the agency cost of debt. *Journal of Financial Economics* 68, 263–285.

Bennedsen, M., F. Pérez-González, and D. Wolfenson, 2010. The governance of family firms, In Baker, H. K., and R. Anderson (Eds.), *Corporate Governance: A Synthesis of Theory, Research, and Practice*. Hoboken, NJ: Wiley.

Berle, A.A., and G.C. Means, 1932. *The Modern Corporation and Private Property*. New York: Macmillan.

Berzins, J., Ø. Bøhren, and B. Stacescu, 2018. Shareholder conflicts and dividends. *Review of Finance* 22, forthcoming.

Berzins, J., Ø. Bøhren, and B. Stacescu, 2018a. Dividends and taxes: The moderating role of agency conflicts. Working paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2973551

Berzins, J., Ø. Bøhren, and B. Stacescu, 2018b. The governance and finance of Norwegian family firms: Main characteristics of the population. CCGR Working Paper 1/2018.

Berzins, J., Ø. Bøhren, and B. Stacescu, 2018c. The performance premium of family firms: Evidence from population data. Working paper.

Bøhren, Ø., and R.Ø. Strøm, 2010. Governance and politics: Regulating independence and diversity in the board room. *Journal of Business Finance and Accounting* 37, 1281–1307.

Bøhren, Ø., and N.E. Krosvik, 2013. The economics of minority freezeouts: Evidence from the courtroom. *International Review of Law and Economics* 36. 48–58.

Bøhren, Ø., B. Stacescu, L. Almli, and K. L. Søndergaard, 2018. When does the family govern the family firm? https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3123589

Demsetz, H., and K. Lehn, 1985. The structure of corporate ownership: Causes and consequences. *Journal of Political Economy* 93, 1155–1177.

Eckbo, B.E. (ed.), 2007. *Handbook of Corporate Finance: Empirical Corporate Finance*, Volume 2. Amsterdam: Elsevier.

Edmans, A. and C.G. Holderness, 2017. Blockholders: A survey of theory and evidence. In Hermalin, B.E., and M.S. Weisbach, *The Handbook of the Economics of Corporate Governance*, Volume 1. Amsterdam: Elsevier.

European Commission, 2009. Final Report of the Expert Group. Overview of Family-Business-Relevant Issues: Research, Networks, Policy Measures and Existing Studies. http://ec.europa.eu/enterprise/policies/sme/

Franks, J., C. Mayer, and S. Rossi, 2009. Ownership: Evolution and regulation. *Review of Financial Studies* 22, 4009–4056.

Hermalin, B.E., and M.S. Weisbach, 2017. *The Handbook of the Economics of Corporate Governance*, Volume 1. Amsterdam: Elsevier.

Jensen, M.C., and W.H. Meckling, 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3, 305–360.

Khanna, T., and Y. Yafeh, 2007. Business groups in emerging markets: Paragons or parasites? *Journal of Economic Literature* 45, 331–372.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny, 2000. Investor protection and corporate governance. *Journal of Financial Economics* 58, 3–27.

Leland, H.E., and Pyle, D.H., 1977. Informational asymmetries, financial structure, and financial intermediation, *Journal of Finance* 32, 371–387.

Maury, B., 2006. Family ownership and firm performance Empirical evidence from Western European countries. *Journal of Corporate Finance* 12, 321–341.

Mehrotra, V., R. Morck, J. Shim, and Y. Wiwattanakantang, 2013, Adoptive expectations: Rising sons in Japanese family firms. *Journal of Financial Economics* 108, 840–885.

O'Boyle, E.H. Jr., J.M. Pollack, and M.W. Rutherford, 2012. Exploring the relation between family involvement and firms' financial performance: A meta-analysis of main and moderator effects. *Journal of Business Venturing* 27, 1–18.

Roe, M.J., 1994. *Strong managers, weak owners: The political roots of American corporate finance*. Princeton, NJ: Princeton University Press.

Shleifer, A., and R.W. Vishny, 1997. A survey of corporate governance. *Journal of Finance* 52, 737–783.

Sraer, D., and D. Thesmar, 2007. Performance and behavior of family firms: Evidence from the French stock market. *Journal of the European Economic Association* 5, 709–751.

Tirole, J., 2001. Corporate governance. *Econometrica* 69, 1–35.

Villalonga, B., and R. Amit, 2006. How do family ownership, control, and management affect firm value? *Journal of Financial Economics* 80, 385–417.

DOI: 10.18261/9788215031583-2018-09

Chapter 8
Legal challenges of state ownership

TORE BRÅTHEN

ABSTRACT The State is not like any other owner. No other owner both owns firms and legislates. Notwithstanding great improvements in corporate governance and the Norwegian Parliament's white papers over the years, substantial legal issues still remain unclear. Role overlap, financial liability, exercise of control over firm's board, management, actions, and structure still remain unresolved.

KEYWORDS: State ownership | law | corporate governance

**All translations are by the author.*

8.1 THE DIFFERENT ROLES OF THE STATE

As a significant shareholder in six of the largest companies on Oslo Stock Exchange, the Norwegian state has the same rights and obligations as other shareholders in similar positions.[1] However, there are several differences between the State and other shareholders in listed companies.[2] The State's primary responsibilities are policy development, legislation and supervision. However, the State's extensive financial resources represent considerable market power, and the State as a whole is Norway's largest supplier and purchaser of goods and services. The State thus plays a whole range of different roles in addition to that of owner.

Since the late 1990s in particular, there has been a strong focus on potential challenges linked to unclear delineation of the State's ownership and other

1. The companies are DNB ASA, Kongsberg Gruppen ASA, Norsk Hydro ASA, Equinor ASA, Telenor ASA, and Yara International ASA.
2. See for example St.meld. nr. 22 (2001–2002) ("*Eierskapsmeldingen*") [the "*White Paper on Ownership*"] and St.meld. nr. 17 (2002–2003), NOU 2004:7 pages 18–19 .

roles.[3] These challenges have been addressed by means of legal and organisational measures.

A guiding principle has been that performance of the State's various roles should, wherever possible, be undertaken by bodies with high specialist expertise, and that the role of owner should thus be separated from the official bodies that formulate sectoral policy or are responsible for regulatory supervision of the area. Accordingly, ownership of companies with no sectoral policy responsibility has been gathered under the Ministry of Trade and Industry since the early 2000s.[4] The "Ownership Department" was made a separate entity in 2001. It concentrates exclusively on ownership issues, although a number of other ministries also have some ownership responsibilities. For example, the State's stake in Equinor ASA is administered by the Ministry of Petroleum and Energy. Even so, the Ownership Department plays a coordinating role.[5] An important consideration underpinning this organisational structure was the desire to separate constitutional responsibility for performance of the ownership role from corresponding responsibility for official and supervisory functions.[6] This approach counters the potential conflicts of interest that may arise from the State's different roles, and facilitates improvement of the public sector's expertise in the area of active ownership.

Potential undesirable features of this combination of roles can also be ameliorated through transparent exercise of ownership in accordance with established active ownership principles.[7] One related measure is the government's classification of state ownership into four categories based on the State's reasons and objectives for direct state ownership.[8] Category 1 contains companies for which the State has only commercial ownership objectives. With one exception, none of the companies in this category are listed.[9] Category 2 includes companies for which

3. The origins of more recent legal regulation of state ownership can be traced to Official Norwegian Report NOU 1989:5 *En bedre organisert stat* ["*On better organisation of the state*"] ("Hermansenutvalget" [the "Hermansen Commission"]). This report was followed by, among others, NOU 1991:8 *Om lov om statsforetak* ["*On the State-Owned Enterprises Act*"], NOU 2003:34 *Mellom stat og marked* ["On independent organisational forms in the state sector"] and NOU 2004:7 *Statens forretningsmessige eierskap* ["*On the state's commercial ownership activities*"].
4. The ministry has been called "Nærings- og fiskeridepartementet" [the "Ministry of Trade, Industry and Fisheries"] since 2013. See further sections 4.1 and 5.2 of St.meld. nr. 22 (2001–2002). Regarding the ministry's internal organisational structure, see NOU 2004:7, pages 53–54.
5. See Meld. St. 39 (2012-2013), pages 140–141.
6. NOU 2004:7, pages 53–54.
7. See Lie, Myklebust and Norvik, *Staten som kapitalist* [The state as a capitalist] (2014), page 17.
8. Regarding this classification, see Meld. St. 27 (2013–2014), pages 55–56, see also page 2 of the State Ownership Report.
9. The exception is Entra ASA.

the State has commercial ownership objectives and the aim of ensuring that the company's headquarters and related functions remain in Norway.[10] Category 3 is defined as companies for which there are commercial objectives and other specifically defined aims, but where state ownership is based on societal considerations other than ensuring that the headquarters remain in Norway. A common feature of companies in categories 1, 2 and 3 is that they engage in commercial competition with other businesses. The key objective guiding the State's commercial ownership of companies in categories 1 to 3 is to achieve the greatest possible return on invested capital over time. The State's ownership of companies in category 4, on the other hand, is primarily motivated by sectoral policy goals.

Some legal aspects of the State's active ownership of listed "category 2 companies" are discussed below. Since the State's interests may not be purely commercial, ownership of these companies may give rise to particularly complicated issues.

8.2 LEGAL CONSEQUENCES OF EXERCISING STATE OWNERSHIP THROUGH PUBLIC LIMITED LIABILITY COMPANIES

The State exercises its ownership of listed companies through companies that are structured as public limited liability companies and are thus subject to the general legislation applicable to entities of this type. Accordingly, the State holds shares in comprehensively regulated entities that can be traded in a thoroughly regulated market.

State ownership of listed public limited liability companies must be exercised within the framework established by applicable legislation.[11] The primary regulatory provisions are found in the Public Limited Liability Companies Act, the Securities Trading Act, the Securities Trading Regulations, the Stock Exchange Act and the Stock Exchange Regulations, as well as the Norwegian Code of Practice for Corporate Governance (NUES).[12] The individuals companies' articles of

10. According to the *State Ownership Report* 2016, category 2 comprises the companies listed in footnote 1 above as well as Aerospace Industrial Maintenance Norway AS, Aker Kværner Holding AS and Nammo AS. See further https://www.regjeringen.no/contentassets/fdcde06c8da8492a8170a61519ff5edc/eierberetning_2017_web.pdf for an overview of the current company classification.
11. See for example Dokument nr. 7 for 1972, see further NOU 1972:38 and NOU 1989:5, pages 105, 109–110.
12. LOV-1997-06-13-45 (Act relating to public limited liability companies (Public Limited Liability Companies Act)), LOV-2007-06-29-75 (Act relating to securities trading (Securities Trading Act)), FOR-2007-06-29-876 (Regulations relating to the Securities Trading Act (Securities Trading Regulations)), LOV-2007-06-29-74 (Act relating to regulated markets (Stock Exchange Act)), FOR-2007-06-29-875 (Regulations relating to regulated markets (Stock Exchange Regulations)) and the Norwegian Code of Practice for Corporate Governance (NUES) 2014.

association are also relevant. Neither the aforementioned laws nor the regulations contain special provisions for companies in which the State has a stake, or special rules defining how the State should exercise its ownership. The same applies to NUES. The articles of association of the listed category 2 companies contain no special provisions linked to state ownership.

Generally speaking, organising State activities through a company rather than a State administrative body entails pursuing those activities through a separate legal person outside the public administration. The company's equity is not included in treasury funds, and the company's transactions are not subject to the budgetary authority of the Storting (the Norwegian parliament) pursuant to Article 75(d) of the Constitution. Such companies are not subject to the traditional authority to issue instructions that otherwise applies in the public administration, and the companies' status as independent legal persons or enterprises, etc. means that, in principle, they fall outside the scope of the Public Administration Act and Freedom of Information Act.[13] Workers are employed by the individual companies, not the State.

In a public limited liability company, shareholders exercise their authority as owners through the general meeting.[14] The general meeting is the company's supreme authority, and may decide all matters not excluded from the general meeting's jurisdiction by law or the articles of association. In addition, some tasks are expressly assigned to the general meeting, such as the appointment of a board majority, or corporate assembly majority if the company has such a body.[15] Other important decisions to be made by the general meeting concern dividends, executive remuneration principles, amendment of the articles of association, changes in capital structure and mergers, demergers and liquidation of the company.[16] Formally, the shareholders have considerable powers, although these have to be exercised in accordance with the procedural rules applicable to the general meeting, including rules on summoning and conducting general meetings. It is through the general meeting that the shareholders constitute the company's supreme authority.

---

13. LOV-1967-02-10 (Act relating to procedure in cases concerning the public administration (Public Administration Act)) and LOV-2006-05-19-16 (Act relating to the right of access to documents held by public authorities and public undertakings (Freedom of Information Act)). Regarding the area of application of the Public Administration Act and Freedom of Information Act, see for example Graver, *Alminnelig forvaltningsrett* [General administrative law] (4th edition, Oslo 2015), pages 292–294, 320–321.
14. Public Limited Liability Companies Act, section 5-1.
15. Public Limited Liability Companies Act, section 6-3.
16. Public Limited Liability Companies Act, section 8-1, section 6-16a, see also section 5-6 (3), section 5-18, section 10-1, section 12-1, section 13-3, section 14-6 and section 16-1.

The majority requirements applicable to general meeting resolutions vary according to matter type. In most cases, the general meeting makes decisions by ordinary majority vote.[17] In the case of elections and appointments, however, the requirement is for the majority of votes cast. Most resolutions regarding amendment of the articles of association, mergers, demergers or liquidation require a two-thirds majority of votes cast and the represented share capital in order to pass.[18] In a small number of cases, a 90 percent majority is required, while shareholder unanimity is required in exceptional cases.[19]

The State holds a sufficiently large stake in the listed category 2 companies that it always has at least negative control, i.e. power to block amendment of the articles of association, mergers, demergers and liquidation, as well as other resolutions requiring a majority of two-thirds or more.[20] In two companies, the State's influence is even greater thanks to its ownership of more than half the shares.[21] In one instance, the State is in an even stronger position, as it owns more than two-thirds of the shares.[22]

Despite its dominant position, the State may not exercise its ownership of listed public limited liability companies without taking into account minority interests that are protected by law. The purpose of the minority-protection provisions of the Public Limited Liability Companies Act is to give minority shareholders the opportunity to safeguard their interests, and to protect minority rights against potential infringement by the majority. The primary provision on minority protection is section 5-21 of the Public Limited Liability Companies Act, which prohibits the general meeting from making any decision "that is likely to give certain shareholders or others an unreasonable advantage at the expense of other shareholders or the company". One reason why this may be relevant to the State is that state ownership may be motivated by non-commercial objectives.[23] The State may thus have different aims from other company shareholders, and may wish to exert its influence contrary to the interests of those shareholders. Such situations

17. Public Limited Liability Companies Act, section 5-17.
18. Public Limited Liability Companies Act, section 5-18, see also sections 13-3(2), 14-6(1) and 16-1(1).
19. Public Limited Liability Companies Act, section 5-19 and section 5-20.
20. According to the State Ownership Report 2016, the relevant ownership interests were DNB ASA (34%), Norsk Hydro ASA (34.26%) and Yara International ASA (36.21%).
21. According to the State Ownership Report 2016, the relevant ownership interests were Kongsberg Gruppen ASA (50.001%) and Telenor ASA (53.97%).
22. According to the State Ownership Report 2016, the ownership interest in Statoil ASA [now Equinor ASA] totalled 67%.
23. See NOU 2004:7, page 32.

typically arise when the State uses its power as owner to turn the company into an instrument for pursuing political aims at the expense of the company's profit objective.[24]

Although the State, through the general meeting, in formal terms may play an active role in the governance of a company, the statutory company structure must still be respected. The State may not exercise its ownership powers in a way that effectively neutralises the board of directors.[25] The Ministry of Trade and Industry has adopted instructions on the management of state ownership interests in limited liability companies which emphasise that the ministry shall not involve itself in decisions falling under the authority of the board and general manager, unless a matter must be considered by the general meeting as a matter of principle. Generally, the ministry's involvement is likely to be restricted to overall strategic governance issues, rather than detailed management of the company's affairs.

8.3 GENERAL COMMENTS ON THE STATE'S PRINCIPLES OF CORPORATE GOVERNANCE

The State launched 10 principles for good governance in 2002 (The State's principles of corporate governance).[26] The original principles were subsequently revised in the Solberg Government's white paper Meld. St. 27 (2013–2014)—the "White Paper on Ownership". The principles specify how the State will act in its capacity as owner, and the State's expectations of companies.[27] According to the White Paper on Ownership, the principles have "created foreseeability regarding

24. It has been claimed that this issue arose when A-pressen, in which Telenor and the Norwegian Confederation of Trade Unions (LO) each held a 48% stake, sold TV2 to the Danish company Egmont contrary to the Minister of Trade and Industry's wishes. The following is stated on page 119 of Lie, Myklebust and Norvik, *Staten som kapitalist* [The state as a capitalist] in this regard: "In this case, the question of principle was again that an attempt was made to exercise the state's influence indirectly, through pressure on the board of directors. It appears clear that, in reality, an attempt was made to use Telenor as a cultural policy instrument, to keep TV2 in the country. This was hardly in the interests of the minority shareholders."
25. See Aarbakke et al., *Aksjeloven og allmennaksjeloven* ["*The Limited Liability Companies Act and the Public Limited Liability Companies Act*"] (4th edition, Oslo 2017), page 448, see also page 329.
26. See St.meld. nr. 22 (2001–2002) on smaller and better state ownership, sections 5.4 and 5.5. See further the *OECD Guidelines on Corporate Governance of State-Owned Enterprises* (2015).
27. See Meld. St. 27 (2013–2014), page 66, see further St.meld. nr. 22 (2001–2002), section 5.4. See also NOU 2004:7, page 51, which states that, "The principles may be regarded as strategies or instruments for achieving the objectives for the administration" of companies in which the state has an ownership interest.

the State's active ownership which has been positively received by stakeholders in the Norwegian capital market".[28]

The current ownership principles for good governance cover various matters: equal treatment of shareholders; transparency about both the State's active ownership and company activities; that owner decisions and resolutions must be adopted at general meetings; that the board is responsible for the development of clear objectives and company strategies while the State has expectations as to company performance; capital structure; board composition; the board's overall responsibility for company management; the board's plan for its own work; skills development and evaluation; pay and incentive schemes; and that companies must make targeted efforts to meet their social responsibilities.[29]

The legal consequences of The State's principles of corporate governance are not always entirely clear. The principles are directed partly at the State in its capacity as owner and partly at the affected companies.

The principle prescribing equal treatment of shareholders is problematic from a purely legal perspective. It is true that both the Public Limited Liability Companies Act and the Securities Trading Act are based on an equality principle, which appears to enjoy widespread support.[30] However, given that the principle requiring equal treatment of shareholders has been made the first ownership principle, it can be questioned whether the State has thereby introduced particular restrictions on its exercise of ownership powers. A phrase in the first White Paper on Ownership may indicate this: "…The State shall also consider the interests of other shareholders where the State may adopt binding resolutions at a general meeting by virtue of its dominant ownership position…".[31] The question is whether minority shareholders may expect even greater respect for the equal treatment principle than in other limited liability companies and public limited liability companies, with the consequence that the State's ownership responsibility may be triggered more easily; see further section 8.4.

To the extent that The State's principles of corporate governance are addressed to relevant companies, they have certain similarities with legally binding instructions to company boards. However, the companies have not adopted The State's principles of corporate governance at a general meeting, and it is therefore difficult to see that board members can be held responsible for non-compliance. How-

28. Meld. St. 27 (2013–2014), page 66.
29. See the overview in Meld. St. 27 (2013–2014), page 67.
30. Public Limited Liability Companies Act, section 5-21; Securities Trading Act, section 5-14.
31. St.meld. nr. 22 (2001–2002), section 5.5.1. However, the phrase is not found in Meld. St. 27 (2013–2014), see particularly page 67.

ever, it does not appear entirely impossible that The State's principles of corporate governance must nevertheless be regarded as a type of instruction to company boards, not least because the State emphasises that companies are expected to comply with them. This raises the question of whether breach of the principles could give rise to liability in damages on the part of the State.

8.4 POTENTIAL LIABILITY IN DAMAGES

The potential liability of the State vis-à-vis company creditors and the company itself is an issue that has played a surprisingly modest role in discussions of the organisation of state ownership. However, the White Paper on Ownership issued by the second Stoltenberg Government—*An Active and Long-Term State Ownership*—provides an example of a focus on liability in damages, in its discussion of liability in connection with instructions from an owner to the board of directors. The white paper states that if a shareholder goes "too far in controlling the company in commercial matters, this may result in creditors filing claims against the State by invoking law of tort or of corporate law concerning piercing of the corporate veil".[32] This statement must be considered in conjunction with the fact that the statutory company structure must be respected; see section 8.2.

Gudmund Knudsen and Sven Ole Fagernæs's 2017 report on ministerial administration of state ownership expands on this view, emphasising for example that both instructions and reversal of board decisions may in principle trigger liability on the part of a shareholder:[33] "If the minister instructs the board through the general meeting or reverses a board resolution, the State assumes responsibility for the decision which, depending on the circumstances, may cause the State to incur liability in damages vis-à-vis the company or others who suffer a loss as a result of the instruction… If the ministry goes too far in controlling the company in commercial matters, this may result in creditors filing claims against the State by invoking law of tort or of corporate law concerning piercing of the corporate veil … State liability pursuant to the provision [in section 17-1(1) of the Limited Liability Companies Act] may be based on the fact that the State has issued instructions at a general meeting that are contrary to law or the company's articles of

32. St.meld. nr. 13 (2006–2007), page 18. See also page 61.
33. Gudmund Knudsen and Sven Ole Fagernæs, *Statsrådens forvaltning av statens eierskap i selskaper som staten eier alene eller er deleier i. Forholdet til Stortinget og selskapets ledelse* ["*Ministerial administration of state ownership of companies in which the state is the sole or a co-owner. Relationship with the Storting and company management*"] (2017), pages 38–39.

association, for example due to infringing the equality principle or the interests of the community of shareholders … but the board implements these nonetheless."

In principle, informal governance signals outside the general meeting may also give rise to liability on the part of a shareholder.[34] In such cases, the legal basis is so-called contributory liability.[35] According to the preparatory works to the act, liability-inducing contribution will typically be deemed to exist "if a shareholder or other person has directly and specifically influenced the particular tortious act of the general manager or relevant person in a position of trust. Relevant influence may include incitement—typically an instruction—which the general manager or person in a position of trust has followed or at least given material weight in his or her deliberations. Relevant contribution may also constitute other specific assistance in undertaking the tortious act".[36] Contributory liability may arise, for example, when a minister makes statements that must in reality be regarded as instructions or incitement.[37] It is an open question whether the combination of specific acts in connection with the appointment of board members (see section 8.5), ownership dialogue (see section 8.6), thorough knowledge of the company's circumstances and specialist expertise in the area may render the State particularly vulnerable in terms of incurring contributory liability.

A further question is whether The State's principles of corporate governance may provide grounds for liability in damages on the part of the State based on the expectations they have created ("justified expectations").[38] For example, may the State incur liability in damages vis-à-vis shareholders if it exercises its ownership contrary to the principle that shareholders should be treated equally? The answer is uncertain.

---

34. Gudmund Knudsen and Sven Ole Fagernæs, *Statsrådens forvaltning av statens eierskap i selskaper som staten eier alene eller er deleier i* [Ministerial administration of state ownership of companies in which the state is the sole or a co-owner], page 48.
35. Public Limited Liability Companies Act, section 17-1(2).
36. Ot.prp. nr. 55 (2005–2006), page 167.
37. See Aarbakke et al., *Aksjeloven og allmennaksjeloven* ["*The Limited Liability Companies Act and the Public Limited Liability Companies Act"*], page 1281, see also 952, stating that both orders and incitement may trigger liability under section 17-1(2) of the Public Limited Liability Companies Act.
38. Regarding liability based on legitimate expectations, see for example Thorson, *Erstatningsrettslig vern for rene formuestap* ["*Protection of purely economic losses under the law of torts*"] (Oslo 2011), particularly page 91 onwards, and Wilhelmsen and Hagland, *Om erstatningsrett* ["*About the law of torts*"] (Oslo 2017), pages 125–128.

8.5 STATE GOVERNANCE THROUGH THE APPOINTMENT OF BOARD MEMBERS

The State does not have its "own" board members in the listed public limited liability companies in which it has a stake, and no special rules apply to board members elected at the suggestion or request of the State. The board members of such companies must be appointed in compliance with the same formal framework as applies to board elections in other public limited liability companies.[39]

State governance of listed public limited liability companies in which the State has an ownership interest has traditionally been effected through the appointment of board members.[40] Over the years, the boards of state-owned companies have been impacted by different roles and role patterns.[41] Emphasis is regularly given to the need for specialist expertise on company boards. Particularly in listed public limited liability companies, where the State must exercise its ownership powers with the interests of other shareholders in mind, there are limits on how much account may be taken of non-commercial considerations.[42] Somewhat varying emphasis has been given to the need for qualifications other than specialist expertise and the ability and willingness to comply with political signals that may challenge purely commercial considerations.[43] The State's principles of corporate governance paragraph 6 states: "Board composition shall be characterised by expertise, capacity and diversity in view of the distinctive characteristics of each individual company."[44]

---

39. See chapter 6 of the Public Limited Liability Companies Act. Like other listed public limited liability companies, companies in which the state owns a stake must additionally comply with the Norwegian Code of Practice for Corporate Governance (NUES) or explain any non-conformances.
40. See Grønlie in Grøndahl and Grønlie (eds.), *Fristillingens grenser* ["*The limits of deregulation*"] (Bergen 1995), page 105 onwards, see also Grønlie, page 118, who writes that, "The state enterprise system was to be "governed without governance".
41. See Grønlie in Grøndahl and Grønlie (eds.): *Fristillingens grenser* ["*The limits of deregulation*"] (Bergen 1995), pages 108–120.
42. See St.meld. nr. 61 (1996–97) Om eierskap i næringslivet [on ownership in the business sector], section 1.2.4: "Our state investments, which are and will remain considerable, must be administered professionally and through elected boards so that no doubt may arise that Norway treats its own and foreign enterprises equally and in accordance with international provisions."
43. St.meld. nr. 13 (2006–2007), page 20, went relatively far in this direction, stating, "Through its representatives on nomination committees, the state will ensure that boards represent a range of expertise and possess sufficient capacity to perform their roles, including that the boards of larger companies include representatives with social understanding and insight." Further, page 47 states that, "The boards shall also lead the companies' strategic efforts. A sound understanding of the company's roles in society and the importance of each individual company for overall industrial development is therefore important".
44. Meld. St. 27 (2013–2014); see the State's ownership principles, paragraph 6.

The appointment of members of parliament and public officials as board members of state-owned companies and other companies in which the State has an ownership interest was long a controversial issue.[45] According to the current guidelines in the Personnel Handbook for State Employees, public officials may not be appointed or nominated if they are employed by a ministry or other central administrative body that regularly deals with matters of material significance to the company, business or industry in question.[46]

The King in Council may make exceptions to the guidelines, but this does not happen in practice.[47] Current practice is that ministers and state secretaries are not elected to the boards or corporate assemblies of companies in which the State is the sole or a co-owner, and that they resign from such positions when they are appointed as a minister or state secretary.

Moreover, since the parliamentary resolution adopted in connection with consideration of Recommendation to the Storting No. 277 (1976–77), members of parliament may not be appointed as members of boards and councils that are subject to the supervisory authority of the Storting.[48] The primary reason for this decision was a fear of mixed roles and protection of the office of board member.[49]

The White Paper on Ownership and the State Ownership Report state that one of the State's most important tasks in its capacity as owner is to help ensure that boards of directors are well-composed and skilled.[50] According to the State Ownership Report, work on board appointments is "a structured process that is ongoing

45. The legal basis for the current rules and practice related to public officials is found in the Storting's consideration of St.meld. nr. 9 (1969–70), see Gudmund Knudsen and Sven Ole Fagernæs, *Statsrådens forvaltning av statens eierskap i selskaper som staten eier alene eller er deleier i* ["*Ministerial administration of state ownership of companies in which the state is the sole or a co-owner*"], page 41. The guidelines were primarily a response to fears of mixed roles and partiality. St.meld. nr. 9 (1969–70) was followed up by means of Innst.S. nr. 91 (1969–70) on the appointment of public officials to boards and councils, etc. At present, guidelines on the appointment of public officials to boards are found in the *Personnel Handbook for State Employees*. See also the former guidelines in St.meld. nr. 40 (1963–64), page 6.
46. See the *Personnel Handbook for State Employees 2016*, section 10.14.1.
47. See Gudmund Knudsen and Sven Ole Fagernæs, *Statsrådens forvaltning av statens eierskap i selskaper som staten eier alene eller er deleier i* [Ministerial administration of state ownership of companies in which the state is the sole or a co-owner], page 41.
48. Innst.S. nr. 277 (1976–77), Recommendation of the Standing Committee on Foreign Policy and Constitutional Affairs regarding the Storting's control of the public administration, page 15 onwards, see also the parliamentary debates at St.forh. 1976–77 Tid.S., page 4074.
49. See Andenæs, *Stortingets kontroll med regjering og forvaltning: Stortingets eget syn* ["*Parliamentary control of the government and public administration: the Storting's own view*"], Jussens Venner volume 1/2 1978, page 1 onwards, particularly pages 15–16.
50. Meld. St. 27 (2013–2014), page 70; State Ownership Report 2016, pages 26 and 28.

throughout the year."[51] In this context, the State evaluates, in its capacity as owner, factors such as board composition, form of working, expertise, efforts, performance, and contribution to value creation. The State also assesses each company's activities and opportunities, the challenges it faces and what expertise the board therefore requires.

The State primarily makes its contribution to the composition of the boards of listed category 2 companies through a nomination committee appointed by the general meeting of each company.[52] One of the nomination committee members is normally an employee of the ministry that administers the State's ownership interest in the company.

Formally, the purpose of the nomination committee is to assist the general meeting and/or corporate assembly. In practice, it plays an important role in evaluation of the board's composition and work by reference to the company's needs, and in making proposals to the general meeting, or in relevant cases the corporate assembly, in connection with election/re-election of board members and board remuneration. The comments on the Norwegian Code of Practice for Corporate Governance (NUES), paragraph 7 state, among other things, that the nomination committee should "consult relevant shareholders to secure candidate nominations and support for the recommendation".[53] It is also common practice for the nomination committee to engage in dialogue with the board or board chair regarding the committee's proposed board candidates.

The members of the nomination committee are representatives of the company. This also applies to members of the nomination committee employed by the relevant ministry. In purely formal terms, therefore, the ministry may not instruct nomination committee members with regard to proposed board candidates, or on other issues related to the performance of their function. In practice, however, members of nomination committees who are employed by ministries charged with administering state ownership have contact with and inform the ministries of committee proposals and ensure that proposals have the support of

51. State Ownership Report 2016, page 28.
52. The Public Limited Liability Companies Act contains no rules on nomination committees, but such rules do exist for financial undertakings whose "total capital under management has exceeded NOK 20 billion for a period of more than 12 months" and which are not subsidiaries in a financial group; see the Financial Institutions Act, section 8-4, see also the Financial Institutions Regulations (FOR-2016-12-09-1502), section 8-5. The Norwegian Code of Practice for Corporate Governance (NUES) (2014), paragraph 7 recommends that a company should have a nomination committee. The procedures of nomination committees are dealt with to some degree in the comments on NUES; see NUES pages 25–26.
53. The Norwegian Code of Practice for Corporate Governance (NUES), page 26.

the minister.[54] The Office of the Auditor General of Norway has pointed out that ministry representatives on nomination committees have in some cases followed instructions issued by political leaders despite personally holding the opinion that the action was inadvisable in view of the company's best interests.[55] In this connection, the Office of the Auditor General has questioned whether the ministry has given adequate consideration to the principle of equal treatment of shareholders.

In 2013, the majority of the Parliament's Standing Committee on Scrutiny and Constitutional Affairs strongly criticised the Minister of Trade and Industry for failing to respect the formal rules in two cases concerning the appointment of board members of listed public limited liability companies involving state ownership. First, the Minister of Trade and Industry was criticised for proposing two new board members for Telenor ASA after the nomination committee had held 23 meetings and settled on seven skilled candidates.[56] Second, the Minister of Trade and Industry was criticised for proposing one new board member (later redesignated as a deputy board member) for Kongsberg Gruppen ASA on the same day as the nomination committee's proposal deadline expired and in a situation where none of the board members was up for re-election. However, the minority of the standing committee found that "not even one breach of The State's principles of corporate governance has emerged. The only negative comment that remains is that on two occasions the Ministry of Trade and Industry gave input on candidates somewhat late, but these members find that this also does not constitute a breach of the guidelines".[57]

Formally, nomination committee proposals related to general meeting/corporate assembly board elections are not binding, and the State's representative at the general meeting is therefore free to vote for board members other than those proposed by the nomination committee, and to vote against the nomination committee's board remuneration proposal. The same applies to the other shareholders. In practice, however, nomination committee proposals are usually followed.

54. See also the description of practice in Dokument 3:2 (2013–2014), pages 39–40.
55. Dokument 3:2 (2013–2014), page 38.
56. Innst. 2009 S—2012–2013, page 8. See also the criticism expressed by the Office of the Auditor General of Norway in Dokument 3:2 (2013–2014), page 41.
57. Innst. 2009 S—2012–2013, page 12.

8.6 STATE GOVERNANCE THROUGH "OWNERSHIP DIALOGUE"

Section 5-1(1) of the Public Limited Liability Companies Act reads:

"The shareholders exercise supreme authority within the company through the general meeting." Neither the State nor other shareholders have any authority over the company other than through the general meeting mechanism. The principle that shareholder decisions and resolutions must be adopted by the general meeting is emphasised in The State's principles of corporate governance paragraph 3.

Neither the Public Limited Liability Companies Act nor the Norwegian Code of Practice for Corporate Governance (NUES) are considered to prohibit contact between shareholders and a company outside the general meeting context.

The White Paper on Ownership discusses how the State, in its capacity as owner, structures its contact with companies.[58] It is stated that a company and its shareholders may exchange information through a variety of channels. In addition to information provided in quarterly and annual reports, other publicly available information and general meetings, regular contact meetings are held with company executives. This is termed the "ownership dialogue", and is regarded as a key aspect of most ministries' company follow-up.[59]

The ownership dialogue allows the State, in its capacity as owner, to gather information about a given company.[60] This is considered to be a key aspect of normal performance of an owner's monitoring and control function. In some more significant matters, such as mergers and demergers, it may additionally be necessary for the owner and company to liaise prior to the related general meeting resolution. The White Paper on Ownership also assumes that the Limited Liability Companies Act and The State's principles of corporate governance do not prohibit the State from raising, in meetings with companies, "matters the companies should consider in connection with their operations and development".[61] Such instances probably come sufficiently close to attempted exertion of control outside the general meeting context that there may be some risk of a misstep. However, in the White Paper on Ownership, the above statement is followed by the

58. Meld. St. 27 (2013–2014), pages 85–86. See also Riksrevisjonens kontroll med forvaltningen av statlige selskaper for 2012 [the Office of the Auditor General of Norway's control report on the administration of state companies in 2012], Dokument 3:2 (2013–2014), page 33, in which the Office of the Auditor General of Norway stated that it had discovered approximately 160 meetings and approximately 160 logged telephone conversations between the Ministry of Trade and Industry and the six companies (Cermaq, DNB, Hydro, Kongsberg Gruppen, Telenor and Yara) for which the ministry was responsible in the period 2010–2012.
59. See for example the State Ownership Report 2016, page 26.
60. St.meld. nr. 22 (2001–2002), section 5.5.3; Meld. St. 27 (2013–2014), page 68.
61. Meld. St. 27 (2013–2014), page 68.

clarification that, "Any view communicated by the State in such a meeting must be regarded as input to the company's administration and board. Matters that require shareholder support must be dealt with at a general meeting."[62] It can be difficult to distinguish between acceptable shareholder contact and governance outside the general meeting context. In a 2015 survey examining shareholder follow-up of social responsibility in companies in which the State has an ownership interest, the Office of the Auditor General of Norway found that ministries "make varying use of the ownership dialogue to challenge companies".[63] The Office of the Auditor General recommended that ministries should "evaluate how they can use the social responsibility-related ownership dialogue to make an even greater contribution to companies' efforts to prevent undesirable incidents".[64] If the State makes statements during such meetings that must be interpreted as governance signals that are difficult for the companies to ignore, there may be grounds for questioning whether the State, in its capacity as owner, can be held liable in damages if the company causes loss to another party through its acts or omissions (see section 8.4).

8.7 CONCLUSION

State ownership of listed public limited liability companies has been a success in many respects.[65] One important reason for this is the gradual development of the legal framework for the exercise of state ownership. There remains, however, several unresolved legal issues, some of which have been discussed in this article. Undoubtedly, more legal research of state ownership of listed public limited liability companies is required.

62. Meld. St. 27 (2013–2014), page 68.
63. Innst. 2006 S—2016–2017, page 11.
64. Innst. 2006 S—2016–2017, page 13.
65. See generally Lie, Myklebust and Norvik, *Staten som kapitalist* [The state as a capitalist] (2014).

DOI: 10.18261/9788215031583-2018-10

Chapter 9
The capitalist state or the state as private owner

SVERRE AUGUST CHRISTENSEN

ABSTRACT This article traces the origins of the considerable state ownership in listed companies in Norway. The Norwegian state is the owner of approximately 30 per cent of the market value of the Oslo stock exchange, and controls companies that account for over half of the market value. The Norwegian parliament has agreed that the state shall operate as a private owner in these companies, respecting other shareholders and the companies' integrity as private enterprises, and thus accommodating the attendant principles of being listed companies. This ownership model for the state developed in the post-war era as a result of the state's ownership in Norsk Hydro; hence, the ownership model is called the Hydro model. The paper will provide a historical explanation of why the Hydro model prevailed, and thereby provide an important explanation for the considerable state ownership in Norway.

KEYWORDS: State ownership | Hydro model | business history | and corporate governance

9.1 INTRODUCTION

A key characteristic of Norwegian businesses and society is the extensive state ownership, particularly in listed companies. The Norwegian state owns approximately 30 per cent of the values in the Oslo Stock Exchange, and controls companies that account for more than half of the market value. This is quite extensive, not least compared with other countries (Storting Report No. 27, 2013–2014). State ownership was a controversial political topic during the post-war era, but broad political support for this ownership developed from the 1980s, along with the belief that it was important to ensure a national anchor for a number of important companies.

From a wider perspective, it may appear paradoxical that state ownership became broadly supported in the 1980s, a decade that was otherwise characterized

by liberalization and globalization. The UK, under its Prime Minister Margaret Thatcher, was at the forefront of this liberalization through the selling and privatization of several state-owned companies. Globalization entailed that an increasing number of companies had foreign owners, and an international division of labor that has particularly benefitted small countries such as Norway. On the other hand, globalization—with international agreements—has weakened the possibility of securing national ownership in businesses. With weak private ownership in Norway, several companies were vulnerable to foreign acquisition. In 2002, the majority in the Storting's (Norwegian parliament's) Standing Committee on Business and Industry stated that the "country must have leading companies with central functions such as headquarters and R&D organizations located in Norway." (Recommendation to the Storting No. 264, 2001–2002). In line with this, state ownership should be considered a form of selective protectionism, i.e. protection against foreign acquisitions (David and Mach, 2003; Christensen 2003).

It is difficult to fully understand the desire for national ownership. Within fields concerned with economics, ownership and economic development, one often finds more or less strong arguments for active, diversified or concentrated, and competent ownership. On the other hand, one would be hard pressed to find a single person from such fields who would argue for why national ownership is important. Nevertheless, one can find many who argue that the value of national ownership is exaggerated (Jakobsen, Goldeng, and Reve, 2001). However, there are several characterizations of ownership that show that most countries are concerned with ensuring national control over particularly important companies (David and Mach, 2003; Christensen, 2003). One expression of this is that while most listed companies in the world can be acquired and sold in what is called the "market for corporate control", this is rarely the case for the most important companies across a range of countries. National ownership in these types of key companies—in Norway often referred to as industry locomotives—was considered important for cooperation between the state and businesses, for innovation, and to ensure a long-term perspective in a variety of ways. More specifically, the desire is to ensure that strategic functions linked to the headquarters, research, and so on, remain in the country.

The concern for national ownership is thus applicable in several countries, so the question is actually why Norway has landed on state ownership as the most important form of selective protectionism. In other countries, concentrated family ownership, combined with other forms of protections against acquisition, has had the same function. According to a global history on corporate governance, capitalism outside the US and UK is "a system where a handful of immensely wealthy

families control almost all of a country's great corporations" (Morck and Steier, 2005, p. 1). For example, in Sweden, the Wallenberg family controls several major companies through shares and preferred shares, thus securing Swedish ownership. In both Germany and Italy, families dominate ownership in large companies such as Fiat and BMW.

The Norwegian historian, Francis Sejersted, believes that the absence of rich and powerful families that could take a nationally leading role after the industrial revolution has characterized Norway since 1814. This can be attributed to the absence of nobility since the Middle Ages, and/or that the financial crises after 1814 hit the upper classes particularly hard. Sejersted believes that the principle of equality was and remains strong in Norway, and that there has been a correspondingly strong skepticism to private power and capital (Sejersted, 1993a, 2005). Recently, Einar Lie has recently presented another theory that there is "a high level of trust in the state as a protector of common interests" (Lie, 2016). Trust in the state—or distrust in private players—could be said to be two sides of the same coin, and both are important characteristics of Norway, which could contribute to explaining the scope of the state ownership.

This contribution emphasizes another cause: the form of the state ownership that developed and that was eventually supported across the political spectrum. But first, a key complaint against state ownership has been that state-owned companies were not sufficiently focused on efficiency and profitability, and that this was due to the lack of zealous private owners who wanted returns on their invested capital. Companies have many stakeholders, with interests invested in different parts, functions and effects of the company. This includes owners, employees, suppliers, customers, local communities, and society at large. An important feature of capitalism is that the owners have the final say in the company. One argument for this is that it is primarily people who have invested their own money who are able and willing to put the profitability requirement ahead of other interests. This could, for example, relate to closing down a part of an enterprise that is losing money even though it affects the interests of employees, suppliers and local communities. It may appear brutal in the individual case, but will also be necessary to ensure societal growth and welfare over time (Bøhren, 2011, p. 198). One objection against the state is that it lacks sufficient capability and willingness to sacrifice the interests of other stakeholders in order to achieve efficiency. One reason for this is that the state, and those who represent the state, lack a pecuniary self-interest. Another is that the state and state companies are vulnerable to pressure from different stakeholders.

The form of state ownership in Norway is referred to as the Hydro model, and is characterized by the state exercising its ownership as a private shareholder, while also respecting minority shareholders. This is to alleviate the traditional problems that ordinary state companies face.[1] The state promises to respect the company's autonomy—and particularly its integrity—as a listed company. The state pledges not to pursue political ends as a company owner, which in practice entails passive ownership. In this respect, the Hydro model means that the state ties itself to the mast as regards governance. The model entails that the companies have private owners, ensuring that profitability is at the top of the agenda. One important element here that has been particularly stressed in recent years is that the companies are public listed companies, so they are subject to market monitoring (Storting Proposition No. 36, 2000–2001, p. 27).

The Hydro model is not the result of a plan or design; it emerged as a result of choices that were made along the way. Nevertheless, the model can be viewed as an attempt to make the best of both worlds: a solid and long-term national foundation, along with the profitability demands of private shareholders and markets. We will see that this was perceived as a successful model, and that this was the basis for this ownership model being chosen for Statoil and Telenor in 2001. These are the two largest companies on the Oslo Stock Exchange, and the state's ownership interests in these companies account for about 80 per cent of the state ownership in the Oslo Stock Exchange. This makes it readily apparent that the Hydro Model is important in explaining the scope of the state's ownership in listed companies.

In the following section, we will follow the state ownership from the Second World War and up to when Telenor and Statoil were listed on the stock exchange in 2001. The primary emphasis will be placed on the background for the general support for the Hydro model from the 1980s. This was particularly related to repeated experiences where companies in which the state was the dominant owner had trouble operating efficiently, and in which the state was left to take responsibility for problems with companies and projects. At the same time, it is not only state-owned companies who have wasted money. In the 1990s, the banking crisis and other scandals tarnished the reputation of private business and industry in Norway, and the state had to save both banks and insurance companies. On one hand, this was important as part of the broad political support that grew for the Hydro model and state ownership. On the other hand, these types of scandals were

1. When the state promises to respect minority shareholders, this is also to avoid accusations of exploitation, which is referred to as "tunnelling", which would then affect the companies' share price.

important in Norway and other countries because this led to more interest in corporate governance, which in turn laid important guidelines for how the state acted as owner.

This work is based on my own research and that of others, and Tore Grønlie's range of work on state ownership; a variety of contributions from Einar Lie on business history in general, and state ownership in particular, should be noted especially. This work attempts to provide a *status* of what the literature can tell us about how the state ended up as such a considerable owner in the Oslo Stock Exchange, and why there is such broad-based support behind this. In conclusion, we will point out key questions for further research.

9.2 STATE OWNERSHIP IN THE POST-WAR ERA

After the Second World War, state ownership in Norway was significantly expanded.[2] The state became an owner in many companies, e.g. within fisheries, electronics and mining. However, three companies formed the center of gravity of the state's ownership: Norsk Jernverk, which the Storting unanimously decided to establish in 1946; Årdal Sunndal Verk (ÅSV), which started by completing the Germans' work on establishing an aluminum plant in Årdal; and Norsk Hydro, which, at 47 per cent, the state became the largest owner in when taking over German shares in the company (Christensen, 1999).

Hydro diverged from the other state companies in that it retained its private identity and method of operation as a public listed company. The state respected Hydro management's overarching goal of generating a profit for their shareholders (Grønlie, 1989; Christensen, 1997). This was not the case for the other state companies. They were not established to make money, but rather to fulfill other objectives such as employment, regional policies, the exploitation of natural resources, and to obtain export revenue. Profitability was a means to achieve these objectives, but not a goal in and of itself. In practice, this meant that other interests superseded the profitability objective. Lie has described how trade unions and local politicians visited decision makers in Oslo to prevent "cost cuts, downsizings or shutdowns" in ÅSV, and thus contributed to "weakening the decision cen-

2. This was due to a number of factors: It was considered necessary to ensure rapid industrialisation; the state took over German ownership in existing and newly established companies. The Labour Party's majority in the Storting provided stable political power, and the war contributed to national and political unity, not least between business leaders and representatives of the Labour Party state. Grønlie 1989; Christensen 1997; Sogner 1994, p. 16.

tres that were working for efficient and rational operations" (Lie, Myklebust, and Norvik 2014, p. 61).

The Labour Party was not as concerned with profitability as private owners, but still wanted the state companies to operate as normal public companies. The party toned down the significance of state ownership, and gave the companies considerable freedom, inviting businessmen to take on roles in the companies and their governing bodies. The Conservative Party, on the other hand, argued in favor of more control of the companies, and in particular control of the state's conduct vis-à-vis the companies (Grøndahl, 1995a; Christensen, 1997). This was in line with the Conservative Party's skepticism regarding the Labour Party state's extensive use of authorizations (Sejersted, 1993b). As regards the view of the Storting's role, there was a healthy dose of opportunism involved—which we can also recognize from our current era. As a government party, the Labour Party wanted a form of governance that would not entail parliamentary responsibility for the Government; as the opposition party, the Conservative Party defended the parliament's primacy and wanted to bring as many issues before the Storting as possible (Christensen, 1997). In practice, however, enterprises that were run well had freedom from state interference, while enterprises that faced and created problems had to accept tighter scrutiny from the state.

A "main approach in Hydro's relationship with the state," according to Øivind Grøndahl, was "the extensive freedom from almost any form of state interference" (Grøndahl, 1995b). Minister of Industry Lars Evensen gave a fundamental clarification early on when he rejected a proposal to instruct the state's representatives on Hydro's board; "I cannot advise a recommendation or order to the state's representatives in a board of this nature" (Christensen, 1997). This would complicate the cooperation with Hydro and businessmen. Moreover, one could "reach further," according to Evensen, "by negotiating directly with the company's administrative and technical leaders" (Christensen, 1997). In addition, the state had considerable influence through frequent contact with Hydro. In 1950, one of the company's board members pointed out "the necessary daily contact Norsk Hydro has with the Norwegian government" (Christensen, 1997). This contact dealt with capital requirements, lease of hydropower, research, the agricultural need for inorganic fertilizer, sensitive products such as heavy water and magnesium, and later on, oil. As an owner, the state was passive, but was active in the relationship with Hydro. The ownership laid important guidelines for the policies the state enacted with regard to Hydro, not least through the fact that state ownership gave the company a legitimacy that paved the way for a policy of expansion (Christensen, 1997, 2003).

The 1960s were characterized by several difficult issues related to state companies. The financial overruns linked to Koksverket became a symbol of a mismanaged state company. A more serious example was the Kings Bay accident in 1962 in which 21 people died. This was the fourth mining accident on Svalbard after the war; a total of 64 people had lost their lives. In addition, ÅSV had recurring problems with polluting spills, and Jernverket struggled with major deficits and bleak prospects. The scandals tarnished the reputation of state industry and contributed to the non-socialist election victory in 1965. However, this did not result in more control of the state companies. "The lessons from Kings Bay and the 'industrial scandals' in 1963/64", according to Tore Grønlie, "appear for all parties to have been to 'keep your hands off'" (Grønlie, 1995). Under the Conservative Party's Minister of Industry, Sverre Walter Rostoft, the companies were given the same amount of freedom as under the Labour Party. He had faith that experienced business leaders would do a good job on the boards and in managing the companies. (Grønlie, 1995).

The Conservative Party's resistance toward state ownership was moderate in the years following the war, and mainly dealt with criticism concerning the organization of state ownership. One important exception was the ownership in Norsk Hydro; the party argued that the state should sell some of its shares—if not all. This applied both when the state became the owner after the war, and became an even bigger issue in the debates on whether the state should fully subscribe in Hydro's issues of shares in 1956 and 1963 (Christensen, 2003). Nevertheless, it was in the 1970s that the confrontation between the Conservative Party and Labour Party regarding state ownership became more pronounced.

9.3 THE 1970S: AMBITIOUS POLITICS AND OVERRUNS

After oil was discovered in the North Sea in 1969, the Conservative government wanted to give Norsk Hydro an advanced role in the petroleum activities, and the state therefore increased its ownership in the company to 51 per cent in 1971 (Storting Proposition No. 63, 1971–72). The Conservative Party hoped that this would prevent the establishment of a wholly-owned state oil company, but the Labour Party took over the government the same year, and Statoil (now Equinor) was established in June 1972 (Storting Proposition No. 113, 1971–72). It is difficult to determine why the Conservative Party supported its establishment; much indicates that they did not want it to become an operative oil company (Aven, 2014). The Labour Party, on the other hand, wanted to develop a strongly integrated Norwegian company that could match the advantage that multinational oil

companies had through their industrial and technological insight. Moreover, they wanted to use Statoil as a tool of industrial policy and, along with the technology agreements, Statoil became important for aiding the development of a Norwegian supplier industry (Christensen and Rinde, 2009). The conflict regarding the Storting's role became relevant as the Labour Party wanted to give Statoil considerable freedom, while the Conservative Party requested "measures that could ensure parliamentary governance and control" (Grønlie, 2001). The Conservatives' desire for control did not diminish from the fact that several Labour Party veterans—Jens Chr. Hauge, Finn Lied and Arve Johnsen—dominated Statoil for the first 10–15 years, the first two as chairs of the board, and Johnsen as CEO.

The state companies would not only create jobs and export revenues; they would also serve as locomotives for industrial and technological development. This particularly applied to Statoil and Kongsberg Våpenfabrikk. The EC referendum, the oil crisis and subsequently the economic crisis in the 1970s, created "an ideal climate for ambitious industry politicians," according to Grønlie (1995c). The phenomenon was far from uniquely Norwegian; several European governments—for example, the Swedish government—practiced an ambitious industrial policy with a strong element of state ownership (Giovanni and Foreman-Peck, 1999; Bohlin, 2014, p. 128; Benner, 1997). In Norway, this development had an additional twist due to the faith in future oil money, and the fact that the Labour Party wanted to stop the progress of the Socialist Left Party. The state therefore spearheaded multiple industrial projects with the aim of increased refinement and value creation in Norway.

Firstly, the state was an initiator in ÅSV increasing the further processing of its own metals, which led to major losses, and also caused the company Alcan to sell all the shares it had purchased in ÅSV in the 1960s (Lie, 2005). Nye Tofte was another ambitious project to produce cellulose sulfate for paper production in order to replace polluting cellulose sulfite. When the project was about to fail due to uncertain economic calculations, the Association of Paper Workers convinced the state to join the project.[3] This did not turn out well, but rather ended in bankruptcy court. In the so-called Emden project, state-owned Sydvaranger contributed 75 per cent of the share capital to the company that would produce sponge iron using Norwegian iron ore and North Sea gas. The project was proceeding as planned until a failing market for sponge iron and higher gas prices spoiled these plans, and the company went bankrupt (Byrkjeland and Langeland, 2000). The official commission that investigated both Nye Tofte and Emden concluded that

3. Multiple wood processing companies were involved, and there were therefore ambitions for the project to contribute to a rationalisation of the industry (Byrkjeland and Langeland, 2000).

the state had not conducted sufficiently thorough commercial risk and profitability assessments (NOU 1983:31). Jernverket was one of several (state) companies that struggled with major deficits, and needed transfers from the state (Fossen, 2013).

The state companies were one thing; another was that the state practiced active politics toward the rest of industry. One example was the extensive support for the shipbuilding industry, which was criticized toward the end of the 1970s for draining the public purse and preventing adjustments in the sector (Lie, 2012). The state also invested heavily in the Norwegian electronics or IT industry in the 1970s, but the majority of the companies underwent crises and/or bankruptcies (Christensen, 2006). The point here is not to discuss whether the policies had favorable consequences. For example, Hans K. Mjelva argues that the shipbuilding support was beneficial because it kept shipyards running, which could later adapt into becoming suppliers for the oil industry (Mjelva, 2005). Neither is the point to argue against state involvement in innovation and business policies. For example, a lot of valuable knowledge emerged from the investment in Norwegian IT industry. There are extensive studies on the significance of national policies for business development. Mariana Mazzucato has received a lot of attention for showing the state's significance for innovation (Mazzucato, 2013).

Generally, however, we can say that politics were more characterized by bold industrial ambitions, but not by corresponding financial governance and control, and/or market orientation. With the exception of companies linked to the oil industry, there are not many Norwegian companies with roots in this effort. A recurring problem with the policy was that different representatives of the state had diverging opinions and priorities. In privately owned companies, it was easier for an owner to enforce the profitability requirement. With the state, there were different representatives saying different things. Moreover, parts of the state apparatus, and the political parties, were vulnerable to pressure from stakeholders. This included pressure from employees, or from powerful industrial players, who were more interested in further refining natural resources than assessing financial risk.

Toward the end of the 1970s, the Labour Party fully acknowledged the problems with the industrial policy and the Government appointed the Lied Committee to lay the foundations for a shift in industrial policy. The committee was headed by the Labour Party's former Minister of Industry, Finn Lied, and he was joined by the head of LO (Norwegian Confederation of Trade Unions), Tor Halvorsen, and former president of the Federation of Norwegian Industries, Jens-Halvard Bratz.[4] The committee claimed that the industrial policy had to be market-based,

4. Jens-Halvard Bratz became Minister of Industry in Willoch's first government.

and that "private profitability coincides with socioeconomic profitability" (NOU 1979: 35). Over the course of the 1980s, "[the Labour Party] moved in a liberal economic direction, where it was no longer a goal for the state to own industry", writes Elin Fossen, who goes on to write: "The primary focus was that business and industry should be profitable, and there were also warnings against public subsidy schemes" (Fossen, 2013, p. 26).

9.4 THE 1980S: LIBERALIZATION AND SCANDALS

One reason for the industrial policy and state companies' challenges in the 1970s was the economic crisis that struck the entire western world. That same crisis laid the foundation for a right-wing wave with less faith in state initiatives and more faith in the market. In Norway, this was manifested through the non-socialist election victory in 1981 that made Kåre Willoch prime minister. He took several steps to phase out and privatize state companies. Here, we will focus on Willoch's skepticism regarding Statoil, which he referred to in his memoirs as "a combination of business, administration, political agency and propaganda machine" (Willoch, 2002; Aven 2014). He was fully supported by his eager state secretary, Terje Osmundsen, who in 1981 wrote the critical book: *Gjøkungen – skal Statoil styre Norge?* (The cuckoo—should Statoil run Norway?)

One objection was that Statoil, through its knowledge and its administrative duties, was given the power to manipulate political processes (Aven, 2014; Lie, 2005). Another was that Statoil's favorable financial terms weakened the company's incentives to streamline and cut costs.[5] Partial privatization of Statoil was never on the table at this point, but Willoch's warnings against making Statoil too powerful were heard. Following negotiations with the Labour Party in 1984, it was agreed that administrative duties would be removed from Statoil and to move large block interests out of the company and over to the State Direct Financial Interest (SDFI). Referred to as the clipping of Statoil's wings, the result was that Statoil became less of a state agency, more resembling an ordinary company.

While Statoil had been favored with the Labour Party's attention and goodwill, Willoch and the Conservative Party developed close ties to Norsk Hydro, and the company became an important contributor to the design of the party's oil policy platform (Aven, 2014; Lie, 2005). To the Conservative Party, it was important to break Statoil's monopoly as a Norwegian oil company. This was the underlying

5. Through significant ownership interests in the commercial oil blocks, the company was destined to achieve major profits.

idea when Hydro was awarded operatorship of the Sølvfaks field, later renamed the Oseberg field, which made the company a bona fide oil company (Storting Proposition No. 109, 1983–84; Recommendation to the Storting No. 290, 1983–84). Hydro was praised for how it solved the operatorship role, not least with important innovations that increased the amount of oil extracted from the wells. Hydro's takeover of ÅSV in 1986 was another important step in the industrial policy. A merger had been in the works for a long time. The Labour Party had wanted this merger to take place for decades, to no avail. However, Hydro rejected a merger, and demanded a takeover of ÅSV—and more importantly here, that the state's ownership should continue to follow the principles of the Hydro model (Lie, 2005). The Conservative Party had no issues accepting these conditions—rather the opposite, and Hydro's takeover was presented as a privatization of ÅSV.

The wing-clipping weakened Statoil, but the major overruns at the Mongstad refinery proved even more significant. It was indicative to many that Hydro withdrew from the Mongstad project in 1986, fearing major overruns. The wholly-owned Statoil, however, continued investing heavily, which caused the company to lose considerable amounts of money, damaging its reputation. The overruns led to the resignation of the board and the company's powerful leader, Arve Johnsen, in 1987. This was a significant blow to Statoil in particular and state operations in general. Willoch used the opportunity to propose that the state should partially privatize Statoil, based on the template of the Hydro model (Willoch, 1987). The Labour Party was far from ready to accept such a proposal, and was instead concerned with the Conservative Party's responsibility for the Mongstad scandal and the issues that were piling up at Kongsberg (Førde, 1987). For Willoch and others, this illustrated the main problem with state operation: firstly, that the companies had no owners, i.e. that there was no involvement of private shareholders, who would have been terrified of damaging overruns; second, that it was futile to believe that the state and politicians could play the role of shareholder.

Mongstad was important, but the truly significant banner cause for the Labour Party, and one which ended up symbolizing the party's change with regard to industrial policy and state ownership, was the crisis in Kongsberg Våpenfabrikk (KV). The company experienced strong growth through the 1970s, increasing sales from NOK 200 million in 1969 to NOK 2.4 billion in 1985 (NOU 1989:2). The previously mentioned Hauge and Lied played key roles, and the company had many accomplishments within technology and products, but profitability was poor. In October 1986, KV requested NOK 600 million in fresh state capital, accompanied by the promise that this would be enough to get the company on a solid financial footing. However, after the oil price drop in 1986, money grew

tighter. In addition, there were many who questioned whether the transfers would put KV on an even keel. One of the skeptics was the company's shop steward, and later LO head, Roar Flåthen (Øyangen, 2014).

Eventually, Minister of Industry Finn Kristensen lost faith in KV's management. The head and board were therefore replaced by a board consisting solely of business people, apart from employee representatives. The new chair of the board Karl Glad set the tone: he and Kristensen fully agreed that KV needed change and, not least, an orientation toward profitability. A highly symbolic meeting took place when the old industry strategists Hauge and Lied met with Prime Minister Gro Harlem Brundtland and Kristensen. "True to form, Hauge offered up a long monologue which emphasized the historical dimensions as well as the industry and defense policy sides of the issue", writes Knut Øyangen in the history of KV, "Finally, the Prime Minister leaned over to look at the Minister of Industry's watch—the time of the old guard was literally over." (Øyangen, 2014).

KV did not go bankrupt; in 1987, a composition solution was determined to be the path forward. This was dramatic for KV, but most people believe it was for the best that there was a growth through the crisis in Kongsberg after the change in approach (Øyangen, Sogner, and Petersen, 2014). It was perhaps equally important as a signal to other state-owned industries that one could no longer ask the state to cover deficits in the companies. This became particularly important during the years around 1990. An underlying point, which was challenging for the Labour party to articulate clearly, was that large sections of industry faced extensive and painful restructuring processes. In Norway, nearly 120,000 industrial jobs disappeared between 1975 and 1990; in Oslo alone, 13,000 industrial jobs disappeared during the financial crisis of 1987–1992 (Statistics Norway, 2003). There was no way to bypass this adjustment; as an example, thousands of industry jobs disappeared as a result of the digitalization of the telecommunications industry. It would be challenging for the state to be left with the responsibility for this. This was most likely an important reason why the Labour Party wanted to scale down its involvement in industry policy and to leave adjustment and restructuring to the market.

The Conservative Party and representatives from businesses had a more principled argument against state operation and for the value of putting private owners in charge of state-owned companies, especially in the context of the financial crisis at the end of the 1980s. The Conservative Party accepted that the state would remain a major owner in Norwegian business and industry, and thus that the Hydro model should be the norm. Willoch talked about the model as a suitable vaccine "against the disease that, based on experience, afflicts pure state companies"

(Willoch, 1986). "I believe in the Hydro model," said the Conservative Party's Else Bugge Fougner in a comment on the major losses in state companies (Aftenposten, 1987). In 1988, Per Kristian Foss said about Statoil that "the Conservative Party wants to gradually develop the state oil company into a Hydro model with mixed state and private ownership" (Foss, 1988).

The Conservative Party's strong resistance to the Hydro model from the postwar era was thus a thing of the past. It was now viewed as a private version of state ownership, and a good and realistic alternative to privatization. In a longer article in *Dagens Næringsliv* in 1990, several people advocated for a partial privatization of Statoil. The company's CEO, Harald Norvik, said that it was too soon, but added that there was "no doubt that Norsk Hydro, at the moment, appears to form a school of thought and point out the final goal for the reorganization of Norwegian state companies" (Dagens Næringsliv, 1990). This was expressed, for example, in the partial privatization of Raufoss Ammunisjonsfabrikker in 1990 (Raufoss Ammunisjonsfabrikker, 2009).

The Labour Party, however, was far from ready for partial privatization of Statoil, and it took some time before the party embraced the Hydro model. On the other hand, the party took the initiative to modernize the state in several ways. The so-called Hermansen Committee's official report on "A better organized state" (NOU 1989:5), was particularly important, and laid the foundation for a general market correction of large parts of the public sector, including "liberalization and deregulation of the energy, telecommunications and post sectors, among others" (NOU 2012:2). This reflected a general skepticism to state governance and bureaucracy, primarily on the grounds that public producers did not receive signals from the market and/or from challenging owners (Arnesen and Hagen, 2008). At the same time, it was important that European integration followed these steps, and that Norway adapted to this. Despite these trends, the development in the 1990s would involve a more critical view of private owners and private business and industry.

9.5 THE 1990S: PRIVATE DOWNTURN, STATE BOOM AND CORPORATE GOVERNANCE

In the 1970s and 80s, crises in state-owned industry dominated the debate on ownership in Norway, and led to the state's retreat. This changed with the banking crisis in the years around 1990, when privately-owned banks squandered billions of kroner on bad—and in part outrageous—banking operations (Lie, 1998; Knutsen, Lange, and Nordvik, 1998). The bank losses are not entirely comparable with the

losses from Mongstad, as an example, because parts of the banks' losses were recouped when the (real estate) market recovered. Nonetheless, the losses in Statoil and KV were modest compared with the total bank losses, which were said to have amounted to NOK 76 billion (Hernes, 2008). In addition, several Norwegian companies lost considerable sums on foreign acquisitions in the 1990s. Kværner's acquisition of Trafalgar House and Saga's acquisition of Santa Fe, both in 1996, did not align with the corporate governance of vigilant private shareholders who were afraid of losing their money. Most notorious was the insurance company Storebrand's attempt to buy the Swedish Skandia, which ended with the company being placed under public administration (Christensen et al., 2017). In many ways, banking and insurance had been the symbol of private capital in Norwegian business and industry, and these industries were now on their knees. This undermined the narrative that if one only allowed private owners, this would be a guarantee against the building of empires and vast financial losses. The narrative about the Hydro model was confirmed by the fact that it became the mode of rescue for failed banks.

In this connection, the Labour Party's Karl Eirik Schjøtt-Pedersen recommended the "Hydro model" for the banks that were now owned by the state after the banking crisis. "The reason is that we do not want hard-handed governance. The banks must be operated commercially" (Aftenposten, 1991). In 1993, the Socialist Left Party's Eilif Meland warned against what would happen when "the EEA Agreement[6] enters into force", because "then these shares will be freely negotiable." A Hydro model could then ensure a national anchoring (Aftenposten, 1993). The state was left with a considerable ownership interest in DnB after the banking crisis; the other large banks in Norway were purchased by foreigners. It was agreed across the political spectrum that the state was necessary to ensure that the largest Norwegian bank, DnB, remained in the hands of Norwegians. The old KV was split into several companies, including Norsk Forsvarsteknologi, which was listed on the stock exchange in 1993 and renamed Kongsberggruppen in 1995. The state was left with 50 per cent of the shares. The Conservative Party's Foss said that he was not "allergic to state equity" and noted that "the Hydro model [was] in fact an excellent way for the state to contribute long-term capital", not least because "the company's performance is constantly measured in the share market through the fact that it is listed on the stock exchange" (Bergens Tidende, 1995).

6. The EEA Agreement is Norway's—along with Iceland and Liechtenstein—agreement with the EU, that provides it with access to the "Internal Market", and which stipulates that Norway must adopt most EU legislation concerning the single market.

In the 1980s, the wing-clipping and Mongstad overruns at Statoil had dominated the news. The news coming from the company became less dire in the 1990s. In fact, there were several people who praised the Norwegian model for oil and gas exploration and management, in which Statoil played a key role (Sejersted, 1997). Not least, Statoil's significance in helping develop the Norwegian supplier industry was emphasized. The fact that the company was run more commercially under Norvik's leadership in the 1990s was noticed, with a more trim and specialized organization (Lerøen, 2002; IAKH, 2016). At the same time, the private company Saga was floundering after its acquisition of Santa Fe. The company was taken over and split between Norsk Hydro and Statoil in 1999. Yet again, a cleanup was needed after a private failure. Saga's shareholders were paid in Hydro shares, which meant that the state's ownership interest in Norsk Hydro was reduced to 43 per cent. In the late 1990s, however, Statoil suffered another misstep with the Åsgard field. The overruns totaled NOK 17 billion, and led to the resignation of the CEO—Norvik—and the board in 1999. At this point, Norvik had already worked intensely for a partial privatization for some time, and the Åsgard overrun became yet another argument for listing the company on the stock exchange (Eger, 2008; Lie et al., 2014; Storsletten 2018).

Before we get to that, we must briefly cover Televerket's development. For the Conservative Party, the agency had been a popular scapegoat throughout the postwar era (Christensen, 2006). Now, the very same organization, without private owners, had gained a completely different reputation. "At the end of the 1980s, we were so pleased with ourselves," said Televerket's Kåre Aarvik, "that the only thing we lacked was the Lord's official blessing. We were living in the clouds, and not entirely without reason." (Aarvik, 1993). Around 1990, Norway had one of the world's best and most efficient and modern telecommunications networks, and the old agency was ready to make investments in other countries (Thue, 2006; Christensen, 2006). The Labour Party's former Minister of Industry, Kjell Holler, who had been at the forefront for the modernization as general director, believed in 1990 that if Televerket was to privatize parts of its operations, a Hydro model was the way to go (Holler, 1990).

In the first half of the 1990s, the discussion regarding Televerket's company structure was important; in 1994 it was renamed Telenor and converted into a public corporation. Grønlie points out that this structure—the public corporation—was a compromise that appeased both supporters and opponents of a strong state. The opponents viewed this as an expression of increased freedom that weakened state influence. Those who wanted a strong state saw this as an adaptation to the prevailing liberalization and globalization, but without relinquishing state control

over central organizations and enterprises in society (Grønlie, 2001). In 1996, Holler's successor, Tormod Hermansen, wanted a partial privatization of Telenor. One of his reasons signaled a new era as regards the view of ownership. He said that with "private owners we would be continuously monitored by players in the international capital market, and that would give us an added pressure that would have a positive impact" (BT, 1996). A few years would pass before Hermansen got his wish. Next we will take a look at one final key trend in the 1990s.

The developments in the 1990s contributed to a more nuanced view of the value of state and private owners. The more positive attitude regarding the state and state ownership is an important value-related part of the broad political support for state ownership. However, this did not mean that the state controlled the companies more; rather the contrary. The principles relating to a need for an arm's length relationship between politics and business were strengthened, and the value of competent and active owners was emphasized by many (Jakobsen et al., 2001). The many scandals in private businesses in Norway and in other countries were actually explained by weak private governance. In the USA, corporate governance had been a topic since the 1970s, but was not widely discussed in the rest of the world until the 1990s (Cheffins, 2015; Christensen et al., 2017). The UK led the way in Europe; several scandals there meant that enterprise management and corporate governance were given high priority, and the "Cadbury Report," among other things, was important (Maclean 1999; Lie et al., 2014). Corporate governance was considered a remedy against managerial capitalism and too-strong administrations. Corporate governance principles were considered particularly important for institutional owners, and these increased in scope. These types of general corporate governance principles have also become normative for the state's exercise of ownership, and are stipulated in the state's "10 principles for good ownership" (Storting Report No. 22, 2001–2002).[7] In other words, this yields less room for the state to influence the companies as owner.

An important part of corporate governance literature is related to the value of the market monitoring that the company is subject to. Market monitoring was emphasized as a significant argument for the partial privatization of Telenor and Statoil in 2000 (Storting Proposition No. 66, 1999–2000; Storting Proposition No. 36, 2000–2001; Lie et al., 2014). Apart from the fact that, presumably, it would lead to more efficient operations, it was also perceived as a democratic benefit that such large and powerful companies were followed more closely by the market, and therefore by the media. This could result in better governance of the compa-

7. They have subsequently been somewhat revised in Storting Report No. 27 (2013–2014).

nies, such as if Statoil was immediately forced to disclose information that was sensitive and relevant to the stock market. This entailed, for example, that if there was a risk of major overruns in projects, the market would be informed. With Mongstad fresh in the memory and major overruns in connection with the development of the Åsgard field in 1999, this was important to Statoil. That point is less convincing now, as Statoil was allegedly late to provide information about the overruns in the Snøhvit field, which took place after it was listed on the stock exchange (Evju, 2008).

The value of active ownership has also been emphasized by many. However, there may be grounds to differentiate between an active industrial and active financial ownership. By industrial ownership, we refer to the fact that the owners have opinions about the operation of the company, such as markets, technology, vertical integration, and so on. By active financial ownership, we refer to the owners stipulating requirements for profitability, returns, capital structure, and transparency. Institutional owners are usually relegated to an active financial ownership; the same usually applies to companies with diversified ownership. Large parts of the corporate governance literature deal with this type of active ownership. In companies with concentrated ownership, one usually sees a combination of industrial and financial ownership. The state is dominant in the companies we are discussing here, so it is a concentrated ownership; nevertheless, the state acts as an institutional owner. It is a firmly established principle that the state shall not interfere with the operation of the companies, but limit itself to an active *financial* ownership. This was, for example, clearly expressed in the proposition that formed the basis for Telenor's listing on the stock exchange, where the state made it clear that it will focus "on questions relating to capital structure, profitability and dividend policy, with an emphasis on long-term profitable business development and value creation for shareholders." (Storting Proposition No. 66, 1999–2000).

Another important trend from the 1990s and beyond was that several players in business and industry came forward to praise and support extensive state ownership, as long as it was conducted in a professional manner. Former chair of the board in Hydro, Einar Kloster, said in 1997 that a long-term and major owner such as the state "gives Hydro considerable stability" (Christensen, 2003). After the many scandals around the new millennium, Statoil's CFO, Inge K. Hansen, claimed that state ownership could be a guarantee against "Enron conditions" because the state had a more long-term perspective (Aftenposten, 2002a). Many have stressed the value of the state as a long-term owner at a time when managers and investors were mostly concerned with the next quarterly figures. Previous

CEO of Statoil, Harald Norvik, stated in 2014 that the "state ownership in Norway is remarkably successful" (Aftenposten, 2014).

Finally, in 2001, the non-socialist parties formed a government with support from the Progress Party, a party that had been a consistent opponent of state ownership since it was established in the 1970s. An important item of the Conservative Party's government platform was the reduction of state ownership. The party wanted to keep a large enough share in most companies to protect against a takeover, but believed that ownership could be significantly reduced in several companies, and expected support from the Progress Party. However, the Progress Party changed its platform in 2002 on the grounds that the Norwegian state had a wealth of money in its Petroleum Fund. In *Aftenposten*, the Progress Party's Christian Tybring-Gjedde wrote that the Conservative Party was stuck in the past: "With a rich state and just a few rich private capitalists, we are choosing to take a new approach" (Tybring-Gjedde, 2003). However, the party's Øystein Hedstrøm believed that it made "no sense for the state to sell out of strategically important Norwegian companies, putting profits of nearly NOK 80 billion in the Petroleum Fund which will then buy minority interests in international companies" (Aftenposten, 2002b). With this, the Progress Party became the final party in the Norwegian political landscape to embrace state ownership and the Hydro model.

Equally importantly in our context, these two arguments have been central to the justification for the extensive state ownership in Norway. To cover the last point first, in other countries a common argument for privatizing state industry was that the state needed money; this is not the case in Norway. Next, Norway lacks private owners who could take over the state's shares. This is an expression of the fact that Norway is a country of modest economic differences. A good picture of this was provided in the book *Staten som kapitalist* (The state as capitalist) from 2014. At that time, the market value of the state's shares in listed companies amounted to about NOK 650 billion. At the same time, if one were to add up the fortunes of "the 100 first entries on (the magazine) *Kapital*'s list of the wealthiest Norwegians," it amounted to about NOK 580 billion (Lie et al., 2014). In other words, a massive transfer of wealth would have to take place if Norwegian capitalists were to take over parts of the state's shares. One of the arguments for reducing the wealth tax in Norway is in fact to create Norwegian ownership centers (Fasting, 2013; Vinje, 2014). One can only speculate whether this type of argument would be better received if state ownership in Norway was perceived as more problematic.

9.6 CONCLUSION

The form of state ownership chosen in Norway, with the state acting as a private shareholder, and with the promise to respect minority shareholders, is important in and of itself. It is also essential as perhaps the most decisive reason for considerable state ownership in Norway today. A significant reason for the support of the Hydro model is that there is a prevailing perception that the model is able to safeguard multiple considerations: a solid, long-term national anchoring to the companies, married with private shareholders' and the market's profitability requirements. In the same way as Grønlie claimed with the public corporations, the model is acceptable to both supporters and opponents of a strong state. Herein lies the background for much of the broad political support of state ownership and the Hydro model. Those who are skeptical to a strong state perceive the model and partial privatization of Telenor and Statoil as the state becoming smaller. Supporters of state involvement see this as a way to ensure that the state plays an important role and retains control.[8]

Although we could indicate this as a partial conclusion—that state ownership is supported for different reasons—we still know too little about how the attitudes regarding state ownership vary over time among interest groups and parties. It is difficult to know what the Labour Party believed and believes about questions of principle relating to how the market economy and capitalism work. One reason is that the party has embraced so many different groupings that there have been different perceptions. Moreover, the party has rarely seen any benefit to making clarifications of its principles. This certainly applied in the 1980s and 1990s, where the party became more positive toward the market, and correspondingly more skeptical toward active state business policy. The Conservative Party, on the other hand, preferred to explain how and why private owners were valuable, and how and why state involvement could be problematic. However, the party is not as clear on why they support state ownership or, put differently, why the party believes that national ownership in certain companies is important. This has contributed to making the debates on state ownership feel like a form of shadow boxing: one has often been more concerned with arguing against the opposing side, instead of arguing for one's own viewpoints. This is even more reason for further research regarding which political motives and incentives lie behind the support of state ownership.

Einar Lies is right that the state enjoys a high level of trust in Norway. At the same time, the very support for the Hydro model is an expression of a lack of trust

8. Grønlie 2001 and Christensen 2003 make the same point.

that the state could run companies efficiently without contributions and corrections from private players and the market. So, it is equally a question of confidence that the state will actually be able to comply with the principles of the Hydro model—or the principles of good corporate governance—in modern terms. An important feature of Norwegian politics today is that the politicians place limitations on their freedom to take action, for example through the fiscal policy rule and corporate governance principles; the fact that the state ties itself to the mast. A more positive angle on this is that the politicians are able to stick to guidelines over time, and that they can be trusted. Although the Storting's Committee on Scrutiny and Constitutional Affairs has contributed to the noise in recent years, an important reason for the support of the state ownership is that the state is generally able to comply with the principles of good ownership. Therefore, further research should take a closer look at the causes of this. This is, for example, important for the question of whether this model is suitable in other countries.

The Labour Party was always a proponent of state companies having a greater degree of freedom, both to avoid parliamentary responsibility, but also to ensure the companies had commercial freedom. It should also be mentioned that the state had considerable influence over the largest companies in the post-war era anyway, through negotiations and agreements on leasing power, licenses, tax conditions, and so on. This secured the state an influence that made it unnecessary to activate the ownership in the companies. This changed from the 1980s. Firstly, the requirement for transparency and equal treatment became more important, which meant that the state could not use leasing power, licenses and regulation to the same extent as policy instruments to influence the companies. Furthermore, the state had relatively less influence in companies such as Hydro, Yara and Telenor, where increasingly large parts of the operations took place abroad. Several people have questioned whether it is right for the state to own companies with such extensive foreign operations. An important question in this context will then be what value these companies have for Norwegian business, industry, and society in general. This would be a natural area for further research.[9]

Finally, state ownership should be placed in the international research as regards it being a de facto expression of a concentrated ownership in the companies, while the exercise of governance is characterized by being a small institutional owner. One thing is that this could yield weak governance of the companies—that the state is not watchful enough. Another is that the state becomes reluctant to have an opinion as owner out of fear of violating the principles of

9. Sverre Herstad has already done some important work in the field (Herstad and Jonsdottir, 2006).

sound corporate governance. If state ownership is linked with state weakness, it could lose its support over time. There should, therefore, be more research on the different preconditions for both the state ownership and for governance.

REFERENCES

Aftenposten (1987, 3 November). Milliardsluket, p. 5.

Aftenposten (1991, 12 November). Staten kan bli "evig" bankeier, p. 10.

Aftenposten (1993, 28 October). Splittelse i Stortinget om aksjesalg, p. 32.

Aftenposten (2002a, 19 August). Staten forsikring mot norsk Enron, p. 27.

Aftenposten, (2002b, 8 June), FrP blokkerer statlig nedsalg, p. 12.

Aftenposten, (2014, 8 November). Det statlige eierskapet i Norge er oppsiktsvekkende vellykket, p. 4.

Arnesen, C. B. and Hagen, K. (2008). *Fra vesen til virksomhet: Et tilbakeblikk på erfaringene med fristilte statlige virksomheter*. (Research report BI Norwegian Business School 3/2008). Retrieved from http://web.bi.no/forskning/papers.nsf/0/7697b088fd5607edc1257458003d6fd2/$file/2008-03-arnesen%20&%20hagen.pdf

Aven, H. (2014). *Høgres syn på statleg eigarskap i norsk oljeverksemd 1970–1984*. (Master thesis in history), University of Oslo, 2014.

Bohlin, J. (2014). Swedish Industrial Policy: From General Policies to Crisis Management, 1950–1980. In C. Grabas and A. Nützenadel (Eds.), *Industrial Policy in Europe after 1945: Wealth, Power and Economic Development in the Cold War* (pp. 113–133). Palgrave Macmillan.

Benner, M. (1997). *The Politics of Growth. Economic Regulation in Sweden 1930–1994* (Thesis). University College of Eskilstuna.

Bergens Tidende (1995, 1 June). Frykter maktkonsentrasjonen, p. 10.

Bergens Tidende (1996, 2 November). Telenor-sjefen ønsker private eiere, p. 8.

Brundtland, G. H. (1997). *Mitt liv – 1939–1986*. Oslo: Gyldendal

Bråthen, T. (2017). *Selskapsrett*. Oslo: Gyldendal Focus.

Bull, E. (1982). *Norgeshistorien etter 1945*. Oslo: Cappelen.

Byrkjeland M. and Langeland, O. (2002). Statlig eierskap i Norge 1945–2000. (Fafo memo 2000:22). Retrieved from www.fafo.no

Bøhren, Ø. (2011). *Eierne, stryet og ledelsen: Corporate governance i Norge*. Oslo: Fagbokforlaget.

Cheffins, B. R. (2015). Corporate Governance since the Managerial Capitalism Era. *Business History Review, 89*(4), 717–744.

Chandler, A. D. (1984). The Emergence of Managerial Capitalism. *The Business History Review*, *58*(4), 473–503.

Christensen, S. A. (1997). *Statens forhold til Norsk Hydro, 1945–1952* (Master thesis in history). University of Oslo.

Christensen, S. A. (1999). Hydro-modellens opprinnelse: Oppgjøret om Norsk Hydro etter andre verdenskrig (Research report, 6). Sandvika: BI Norwegian Business School.

Christensen, S. A. (2003). Statlig eierskap og nasjonal kontroll. In S.A. Christensen, H. Espeli, E. Larssen and K. Sogner (Eds.), *Kapitalistisk demokrati? Norsk næringslivshistorie gjennom 100 år* (pp. 67–148). Oslo: Gyldendal.

Christensen, S.A. (2006). *Switching Relations: The Rise and Fall of the Norwegian Telecom Industry* (PhD-dissertation). Oslo: BI Norwegian School of Management.

Christensen, S.A. and Rinde, H. (2009). *Nasjonale utlendinger: ABB i Norge 1880–2010*. Oslo: Gyldendal.

Christensen, S.A. (2014). Globaliseringens fortellinger: ABB Norges konsernansvar for olje- og gassvirksomhet. In E. Ekberg, M. Lönnborg and C. Myrvang (Eds.), *Næringsliv og historie* (pp. 61–92). Oslo: Pax.

Christensen, S.A., Ekberg, E., Bergh, T. and Myrvang, C. (2017). *Skadeskutt og livskraftig 1945–2017, Storebrand og forsikringsbransjen, bind 2: 1945–2017*. Oslo: Universitetsforlaget.

Dagens Næringsliv (1990, 27 September). Statoil-toppene vil jobbe langsomt for Hydro-modell.

David T. and Mach A. (2003). The Specificity of Corporate Governance in Small States: Institutionalization and Questioning of Ownership Restrictions in Switzerland and Sweden. In M. Federowicz and R.V. Aguilera (Eds), *Corporate Governance in a Changing Economic and Political Environment. Studies in Economic Transition* (pp. 220–246). London: Palgrave Macmillan.

Eger, K. M. L. (2008). *Spillet om Statoil: En casestudie av beslutningsprosessen som endte med delprivatisering av Statoil* (Master thesis in Political Science). University of Oslo.

Espeli, H. (1992). *Industripolitikk på avveie*. Oslo: Ad Notam Gyldendal.

Evju, I. (2008). *Mongstad, Åsgard og Snøhvit: Politisk oppmerksomhet ved budsjettoverskridelser i Statoil* (Master thesis in Political Science). University of Oslo.

Fasting, M. (2013). Statens rikdom hindrer privat eierskap. *Magma*, (1), 76–77.

Fossen, E. (2013). *Statens industripolitikk på 1980-tallet: Mot en mer markedsliberal industripolitikk i Norge* (Master thesis in history). Retrieved from www.duo.uio.no

Foss, P. K. (1988, 28 April). Milliardsløsing uten grenser?, *Aftenposten*.

Førde, E. (1987, 13 October). Den uansvarlege Willoch, *Aftenposten.*

Giovanni, F. and Foreman-Peck, J. (Eds.). (1999). *European Industrial Policy: The Twentieth-century Experience*. Oxford: Oxford University Press.

Grøndahl, N. Ø. and T. Grønlie (Eds.). (1995). *Fristillingens grenser*. Bergen: Fagbokforlaget.

Grøndahl, N. Ø. (1995a). Partienes styringspolitikk. In N.Ø. Grøndahl and T. Grønlie (Eds.), *Fristillingens grenser: statsaksjeselskapet – styringsproblemer og reformprosesser gjennom 50 år* (pp. 85–104). Bergen: Fagbokforlaget.

Grøndahl, N. Ø. (1995b). Tre styringsprofiler Norsk Jernverk, Norsk Hydro og Statoil. In N.Ø. Grøndahl and T. Grønlie (Eds.), *Fristillingens grenser: statsaksjeselskapet – styringsproblemer og reformprosesser gjennom 50 år* (pp. 125–149). Bergen: Fagbokforlaget.

Grønlie, T. (1989). *Statsdrift: Staten som industrieier i Norge 1945–63*. Oslo: Tano.

Grønlie, T. (1992). Establishment of state-owned industrial enterprises: Norway in a West-European context. *Scandinavian Journal of History*, *17*(2–3), 209–225.

Grønlie, T. (1995c). Styrenes styring: et mangfold av styreroller. In N.Ø. Grøndahl and T. Grønlie (Eds.), *Fristillingens grenser: statsaksjeselskapet – styringsproblemer og reformprosesser gjennom 50 år* (pp. 105–124). Bergen: Fagbokforlaget.

Grønlie, T. (2001). Mellom politikk og marked: organisering av statlig næringsdrift. In B.S. Tranøy and Ø. Østerud (Eds.), *Den fragmenterte staten: Reformer, makt og styring* (pp. 301–332). Oslo: Gyldendal.

Hernes, G. (2008, 11 April). Sosialisme for kapitalismens skyld, *Morgenbladet*, p. 24.

Herstad, S. and Jonsdottir, A. (Eds.) (2006). National innovation systems and domestic multinational corporations: Key issues, dynamics and implications (Synthesis report from the DOMUS research project). Retrieved from http://www.diva-portal.org/smash/get/diva2:707105/FULLTEXT01.pdf

Holler, K. (1990, 6 June). Behovet for et sterkt televerk. *Aftenposten*, p. 2.

Recommendation to the Storting No. 91 (1969–70).

Recommendation to the Storting No. 290 (1983–84). *Utbygging av Osebergfeltet: ilandføring av olje til Norge og etablering av en råoljeterminal m.v.*

Recommendation to the Storting No. 264 (2001–2002) *Et mindre og bedre statlig eierskap*.

Recommendation to the Storting No. 209 S (2012–2013).

IAKH, Department of Archaeology, Conservation and History, University of Oslo (2016), Forprosjekt Statoils historie 1972–2022. Retrieved from http://www.hf.uio.no/iakh/forskning/prosjekter/statoils-historie/forprosjekt-statoil-06-04-2016-kr.pdf

Jakobsen, E. W., Goldeng, E. and Reve, T. (2001). Eierskap: spiller det noen rolle? In K. Roland, V. D. Norman and T. Reve (Eds.), *Rikdommens problem: oljeformue, eierskap og fremtidens pensjoner* (pp. 147–171). Oslo: Universitetsforlaget.

Jones G. (2004). *Multinationals and Global Capitalism: From the Nineteenth to the Twenty-first Century.* Oxford: Oxford University Press.

Knutsen, S., Lange, E. and Nordvik, H. W. (1998). *Mellom næringsliv og politikk: Kreditkassen i vekst og kriser: 1918–1998*. Oslo: Universitetsforlaget.

Lerøen, B. V. (2002). *Dråper av svart gull: Statoil 1972–2002*. Stavanger: Statoil.

Lie, E. (1998). *Den norske Creditbank 1982–1990: En storbank i vekst og krise*. Oslo: Universitetsforlaget.

Lie, E. (2001). Langsiktighet eller styringssvikt i den økonomiske politikken? In B.S. Tranøy and Ø. Østerud (Eds.), *Den fragmenterte staten: Reformer, makt og styring* (pp. 190–2016). Oslo: Gyldendal

Lie, E. (2005). *Oljerikdommer og internasjonal ekspansjon: Hydro 1977–2005*. Oslo: Pax.

Lie, E. (2012). *Norsk økonomisk politikk etter 1905*. Oslo: Universitetsforlaget.

Lie, E., Myklebust, E. and Norvik, H. (2014). *Staten som kapitalist: Rikdom og eierskap for det 21. århundre*. Oslo: Pax.

Lie, E. (2016). Context and Contingency: Explaining State Ownership in Norway. *Enterprise & Society*, *17*(4), 904–930.

Lange, E. (1991). Førsteopponentinnlegg til Tore Grønlies doktordisputas 'Statsdrift'. *Historisk tidsskrift, 70*(3), 406–418

Maclean, M. (1999). Corporate Governance in France and the UK: Long-Term Perspectives on Contemporary Institutional Arrangements. *Business History*, *41*(1), 88–116. doi: 10.1080/00076799900000203

Mazzucato, M. (2013). *The Entrepreneurial State.* London: Anthem.

Morck, R. K. and Steier, L. (2005). The global history of corporate governance: An introduction. In R.K. Morck (Ed.), *A history of corporate governance around the world: Family Business Groups to Professional Managers* (pp. 1–64). Chicago: University of Chicago Press.

Storting Report No. 39 (2012–2013). *Mangfold av vinnere: Næringspolitikken mot 2020*, Nærings- og fiskeridepartementet.

Storting Report No. 27 (2013–2014). *Et mangfoldig og verdiskapende eierskap* (Eierskapsmeldingen).

Mjelva, H. K. (2005). Tre storverft i norsk industris finaste stund: Ein komparativ studie av Stord verft, Rosenberg mek. verksted og Fredrikstad mek. verksted 1960–1980, (Rokkansenteret Rapport 6). Bergen: Stein Rokkan Center for Social Studies.

NOU 1979: 35 *Strukturproblemer og vekstmuligheter i norsk industri*.

NOU 1983: 31 *Statens engasjement i Tofte- og Emden-konkursene*.

NOU 1989: 2 *Kongsberg Våpenfabrikk*.

NOU 1989: 5 *En bedre organisert stat*.

NOU 2004: 7 *Statens forretningsmessige eierskap – Organisering og forvaltning av eierskapet.*

NOU 2012: 2 *Outside and Inside – Norway's agreements with the EU.*

Osmundsen, T. (1981). *Gjøkungen: Skal Statoil styre Norge?* Aventura.

Raufoss Ammunisjonsfabrikker. (2009, 14 February). I Store norske leksikon. Retrieved 22 May 2018 from https://snl.no/Raufoss_Ammunisjonsfabrikker

Reich, R. (2007). *Supercapitalism: The Battle for Democracy in an Age of Big Business. New York: Knopf.*

K. Roland, V. D. Norman and T. Reve (Eds.). (2001). *Rikdommens problem: oljeformue, eierskap og fremtidens pensjoner*. Oslo: Universitetsforlaget.

Sejersted, F. (1997). Innledning. In F. Sejersted and O.E. Olsen (Eds.), *Oljevirksomheten som teknologiutviklingsprosjekt* (pp. 11–25) Oslo: ad Notam.

Sejersted, F. (1984). *Høyres historie: Opposisjon og posisjon 1945–1984*. Oslo: Pax.

Sejersted, F. (1993a). Den norske "Sonderweg". In F. Sejersted, *Demokratisk kapitalisme* (pp. 163–208). Oslo: Universitetsforlaget.

Sejersted, F. (1993b). Sosialdemokratiet finner sin form: Kampen om fullmaktslovene etter 1945. In F. Sejersted, *Demokratisk kapitalisme* (pp. 271–302). Oslo: Universitetsforlaget.

Sejersted, F. (2013). *Sosialdemokratiets tidsalder: Norge og Sverige i det 20. århundre*. Oslo: Pax.

Skard, H. (1972). *Franske statsmyndigheter og herredømmet over Norsk Hydro under første verdenskrig* (Master thesis in history). University of Oslo.

Sogner, K. (1994). *Fra plan til marked: Staten og elektronikkindustrien på 1970-tallet* (TMV Skriftserie nr. 9). Oslo: TMV-senteret.

Sogner K. (with contributions from Christensen, S.A.) (2001). *Plankeadel: Kiær- og Solbergfamilien under den 2. industrielle revolusjon*. Oslo: Andresen and Butenschøn for BI Norwegian Business School.

Sogner, K. and Petersen, T. (2014). *Strategiske samspill: Kongsberg Gruppens historie 1987–2014*. Oslo: Pax.

Statens eierberetning 2016.

Storsletten, A. V. (2018) *Prosessen bak delprivatiseringen av Statoil i 2001 – fra selskap til vedtak* (Master thesis in History). University of Oslo.

Storting Report No. 61 (1996–97). *Om eierskap i næringslivet*.

Storting Report No. 22 (2001–2002). *Et mindre og bedre statlig eierskap*.

Storting Report No. 13 (2006–2007). *An active and Long-Term State Ownership*.

Storting Report No. 27 (2013–2014). *Diverse and value-creating ownership*.

Storting Proposition No. 63 (1971–72). *Om kjøp av aksjer i Norsk Hydro*.

Storting Proposition No. 113 (1971–72). *Opprettelsen av Oljedirektoratet og et statlig oljeselskap mv.*

Storting Proposition No. 109 (1983–84). *Utbygging av Osebergfeltet, ilandiøring av olje til Norge og etablering av en råoljeterminal m.v.*

Storting Proposition No. 66 (1999–2000). *Om Telenor.*

Storting Proposition No. 36 (2000–2001). *Eierskap i Statoil og fremtidig forvaltning av SDØE*.

SSB, Statistics Norway (2003). *Industri og bergverk: Sysselsatte i bedrifter, etter fylke. 1966–2000*. Retrieved from https://www.ssb.no/a/kortnavn/indsys/tab01.html

Thue, L. (2006). Nye forbindelser: Norsk telekommunikasjonshistorie 1970–2005. Oslo: Gyldendal.

Tranøy, B. S. and Østerud, Ø. (Eds.). (2001). *Den fragmenterte staten: Reformer, makt og styring.* Oslo: Gyldendal.

Tveite, T. (1995). Forvaltningsselskap for statsindustri? In N.Ø. Grøndahl and T. Grønlie (Eds.), *Fristillingens grenser: statsaksjeselskapet – styringsproblemer og reformprosesser gjennom 50 år* (pp. 151–186). Bergen: Fagbokforlaget.

Tybring-Gjedde, C. (2003, 11 May). Hva skal vi leve av etter oljen? *Aftenposten*, p. 10.

Vinje, V. (2014). *Kompetent kapital: Innovasjon, eierskap og skatt* (Civita report). Retrieved from https://www.civita.no/assets/2013/12/Kompetent-kapital-Innovasjon-eierskap-og-skatt.pdf

Willoch, K. (1986, 27 November). KV til Statoil: en snarvei til subsidier. *Aftenposten*, p. 46.

Willoch, K. (1987, 5 October). Statsbedriften Statoil og Mongstad: Bedrøvelig og kostbar erfaring. *Aftenposten*, p. 2.

Willoch, K. (1990). *Statsminister*. Oslo: Schibsted.

Willoch, K. (2002). *Myter og virkelighet: om begivenheter frem til våre dager med utgangspunkt i perioden 1965–1981*. Oslo: Schibsted.

Øyangen, K. (2014). *Moderniseringslokomotivet: Kongsberg Våpenfabrikks historie 1945–1987.* Oslo: Pax.

Aarvik, K. (1993). *Makt og avmakt: om organisasjon og ledelse i Televerket: fra "væsen" til "verk".* Oslo: Norsk Telemuseum.

DOI: 10.18261/9788215031583-2018-11

Chapter 10
The Economics of Auditor Regulation

JOHN CHRISTIAN LANGLI AND MARLEEN WILLEKENS

ABSTRACT Financial statement auditing is useful as it provides assurance about the reliability of financial information that firms issue. To protect companies' stakeholders and to safeguard audit quality, financial statement auditing is heavily regulated. We explain why and when auditor regulation is necessary and warranted, and discuss important recent changes in auditor regulation, such as mandatory audit firm rotation in the EU. We also identify a number of auditing-related issues that require further deliberation.

KEYWORDS: Auditing | regulation | financial stability

10.1 INTRODUCTION

External (financial statement) auditors provide assurance about the reliability of the financial information that companies issue. This information can be used by shareholders and other stakeholders for decision making—for example, to decide whether to buy or sell shares in the company, or whether a loan should be granted or extended to the firm. It is the management of the company that prepares the company's financial statements. Many stakeholders do not have financial and accounting skills to judge whether the information that management provides is reliable. Moreover, even if they have so, they do not have access to the inside information that corroborates the financial statement information. Company management may have incentives and opportunities to "lie" about the true economic condition of the firm, and to provide financial information that is distorted. External auditors are technical experts that can mitigate this information risk by investigating the financial statements of a company and the underlying transactions, and attesting to whether the financial information that management provides is reliable or not. Financial statement auditors should be independent from the firm

to avoid collusion with management, and their main role is to provide assurance to the users of the financial statements that the latter are reliable.

Financial statement auditing is a *professional*, *economic* and *regulated* activity executed by *individuals* with the help of *audit technology* (Hay, Knechel, and Willekens, 2014). All these aspects of auditing are interrelated and jointly affect what the eventual quality of an audit will be. An audit is conducted by professionals who have acquired the specific skills and knowledge necessary to perform the audit, and who possess the appropriate license to mark themselves out as a professional. In the course of an audit, specialized technology is used to augment the professional expertise of individuals. Further, audits are economic goods in the sense that a market exists to match those who will supply an audit with those who would demand an audit. In that market, auditors compete with each other to obtain new clients. However, audits and the audit market are also very heavily regulated. Who needs an audit, who can supply an audit, and the conditions under which the two parties can contract for audit services are all subject to various forms of regulation. There is also a risk of litigation against the auditor in the case of malpractice. Because auditing is a human activity conducted by individuals, the quality of a specific audit is conditional on individual auditor characteristics, the incentives that auditors face, as well as the audit (and audit- and accounting-related) regulation. Given all these aspects, auditing is a complex phenomenon to study and understand.

There exists a rich academic literature that investigates the various drivers that affect financial statement auditing and audit quality (for reviews, see DeFond and Zhang, 2014; Francis, 2004; Langli and Svanström, 2014; Lennox and Wu, 2017; and Vanstraelen and Schelleman, 2017). In this chapter, we will only zoom in on economic theories that relate to audit regulation and how regulation has developed. Auditing is heavily regulated, and both demand and supply/production of auditing is subject to various forms of regulation. In the past decade in particular, we have seen a large increase in auditor-related regulation globally as various accounting and auditing scandals triggered deeper audit regulation under the premise of improving audit quality. The best-known example is the Sarbanes-Oxley Act in the United States. The latest European audit legislation (European Union, 2014 a and b) is one of the more recent regulatory changes that fits under this umbrella. It is worth noting that the new regulation primarily concerns audits of public interest entities (PIEs) such as banks, insurance companies and listed companies, *and* the audit firms that serve these companies. For non-PIEs and audit firms that only audit non-PIEs, the regulation has hardly changed. And in countries with mandatory auditing for non-listed firms, which is the case in, e.g., the EU and Norway, regulation has been eased as small companies are now allowed to opt out of auditing.

This chapter proceeds as follows: Next, in section 10.2, we discuss the economic theory of audit regulation. Section 10.3 describes important new audit regulations that have been put in place in the aftermath of the financial crisis of 2007–2008. We limit our discussion to the developments in the EU and the United States. In section 10.4, we identify research opportunities related to audit regulation, and point to a severe challenge for auditing researchers, namely their lack of access to relevant data from audit firms and regulatory bodies.

10.2 ECONOMIC THEORY OF AUDIT REGULATION

There are various theories of regulation. In this section, we will only focus on *economic* theories that apply to the audit setting.

10.2.1 WHY IS THERE AUDIT REGULATION?

According to economic theory of regulation, there is a prima facie case[1] for regulatory intervention when there is market failure that is accompanied by private law failure (Ogus 1994). Market failures are usually caused by lack of adequate information, lack of competition or by externalities, and exist when the quantity or quality of a good supplied differs from the socially efficient outcome. In such cases, government regulation that moves the private output of a good closer to the socially efficient solution can improve social welfare (efficiency) in a Pareto sense. Applied to the market for audited financial information, market failure exists if the output of audited information in annual reports or distributed via other corporate communication channels is non-optimal in a Pareto sense in the absence of audit regulation (Eilifsen and Willekens 2008).

An early rationale for audit regulation was to protect the users of financial information (i.e. the investors), as they were presumed not to have perfect access to company information (see, for example, Watts and Zimmerman 1986). Disclosure choices of accounting information might create negative external effects (or externalities[2]) to

1. Note that it is only a prima facie, and not a conclusive, case for such intervention. The reason is that the regulatory solution may be no more successful in correcting the inefficiencies than the market or private law, or that any efficiency gains to which it does give rise may be outweighed by increased transaction costs or misallocations created in other sectors of the economy.
2. Although 'market failure' and 'externality' are related concepts, and sometimes used interchangeably, they are not synonymous. Market failure can be caused by externalities, but can also be caused by other factors, such as monopoly situations, public goods, and informational asymmetry (see, for example, Cooter and Ulen, 1988).

users of this information, for instance by not providing adequate information about bonus plans for top management, or the terms of sale and leaseback transactions. In general, an externality exists where an action of one economic agent affects the utility of another in a way that is not reflected in the market place (Just et al., 1982: 269). Losses to financial statement users due to resource allocation decisions based on defects in *audited* financial statements that the auditor did not detect or report, could be seen as an 'externality'. Directors and the external auditors can thus be considered as *jointly* responsible for 'hazardous' or misleading financial reporting. The externality is aggravated by informational asymmetry. Directors and external auditors have more information about the value of the firm than external parties. In addition, the delivered audit quality by auditors cannot be observed by clients and third party users of financial statements. The social objective of audit regulations and liability could be seen as means to correct for various externalities created by directors *and* external auditors such that total social utility is improved. Policies adopted to correct for hazardous financial reporting behavior—that is, the behavior of the directors—will necessarily have an impact on audit demand, since the directors are the ones who acquire audit services. Policies adopted to correct for externality-generating audit behavior will necessarily affect audit production—or the behavior of the auditor (Willekens, Steele and Miltz, 1996).

10.2.2 DIFFERENT TYPES OF REGULATION IN THE AUDIT CONTEXT

Ex ante regulation and ex-post liability are two very different approaches to control for activities (such as the production of audit services) that create risks of harm to third parties or externalities. From its beginnings, the literature on optimal regulation has focused on alternative types of ex ante policies, such as safety standards, Pigouvian taxes, and transferable permits. Ex ante policy instruments modify behavior in an immediate way through requirements that are imposed *before*, or at least independent of the actual occurrence of harm, and are *public* in nature (Shavell 1984). Ex ante rules can be pronounced directly by the state through laws, or the state can delegate its authority to another body. In the context of audit regulation, an ex ante policy that applies to the auditee is the statutory audit requirement, or the obligation to appoint an external financial statement auditor to attest the reliability of the financial statements. For the auditor, the auditing standards—International Standards on Auditing (ISAs), or local professional auditing standards—could be seen as a form of "ex ante rules." As opposed to disclosure standards—those, for instance, included in the International Financial Reporting

Standards (IFRSs)—auditing standards are rather general in the sense of stressing objectives rather than precise auditor actions to reach those objectives. In addition, auditor independence regulations, such as the prohibition to perform certain non-audit services for audit clients, and the requirement for audit firms to rotate every so many years, are other examples of ex ante standards. Note that by setting very precise ex ante standards, audit standards become less vague at the risk of being mis-specified. For example, by forbidding auditors to perform certain non-auditing services to clients, audit quality need not necessarily be improved; the opposite may even occur. The reason is that audit quality may improve due to spillover effects obtained from providing non-audit services.

The second policy instrument, namely liability in tort, works through the deterrent effect of damage actions that may be brought *after* harm has occurred, and hence is *private* in nature (Shavell 1984). The threat of suit causes the potential injurer to internalize the expected social harm and, hence, to take optimal precaution. In the audit case, this would imply that audit liability is an incentive for the auditor to produce an optimal level of audit quality. It is only since the 1980s that researchers have also analyzed the ability of exposure to ex post liability to correct for externalities.[3] The basic premise of law and economics is that legal rules create implicit prices on behavior, and that the responses of individuals and organizations to those prices can be analyzed in exactly the same way that responses to explicit prices can be analyzed (Ulen 1993). The threat of litigation to audit firms can, thus, be expected to affect audit production behavior because auditors are incentivized to provide audits of sufficiently high quality. The potential liability of corporate management (that is, the directors) to third parties might also affect the demand for auditing services and other monitoring mechanisms.[4] It is, however, important to realize that alternative liability regimes[5] may affect behavior of auditors and auditees differently, and may result in alternative resource alloca-

3. Cooter (1991), for example, points that the elaboration of price theory by mathematical economists took the legal framework for granted. Liability law is, however, an important mechanism for allocating resources. Nowadays, economic theories tend to understand liability law as a search for efficiency in incentives and risk bearing.
4. The "deep pocket" hypothesis states that larger audit firms have greater incentives to provide high quality audits than smaller audit firms because they have more wealth at risk (Lennox 1999; Khurana and Raman 2004). Top management may therefore prefer large audit firms, such as EY, KPMG, Deloitte and PWC (often referred to as Big Four), because better audit quality reduces litigation risks. A large body of literature shows that Big Four audit firms provide audits of higher quality than other audit firms (DeFond and Zhang 2014).
5. Such as strict liability versus a negligence standard, or joint and several liability versus proportionate liability (Naravanan 1993; Schwartz 1998).

tions, some of which might not be socially efficient. A similar remark holds for alternative legal environments, such as common-law systems versus civil law systems. An obvious question is which enforcement mechanisms work as a deterrent against unwanted auditor behavior in environments where ex post liability is less prevalent, as is the case in various European countries.

Note that Willekens et al. (1996) analyze the joint use of ex ante standards and ex post liability in the auditing setting. They also study the impact of uncertainty (or vagueness) about auditor due care (negligence) on audit quality, and the role of the professional auditing standards, Generally Accepted Auditing Standards (GAAS), in such a setting. They show that the vagueness of legal negligence standards can either have a positive or negative effect on audit quality, and that this depends on the level of vagueness (or precision). Relatively little (substantial) vagueness will have a positive (negative) impact on audit quality as compared to the situation where negligence standards are clear. It is shown that clear GAAS accompanying a vague legal negligence standard have a positive impact on audit quality or effort. As the precision of audit regulations and the importance of ex post liability is not constant across countries, the impact of audit regulation and ex post liability on audit behavior can be expected to be very different across the globe. One may then wonder why the new auditor independence rules (see below) are so similar in most countries.

10.3 AUDIT QUALITY AND CONTEMPORARY HOT ISSUES IN AUDIT REGULATION

In this section, we will discuss some key aspects of contemporary audit regulation in Europe. Note that most of these regulations are also important outside Europe, and where relevant we will elaborate on this. In 2006, the European Commission issued its Statutory Audit Directive (European Union 2006) that aimed at a high level of, but not full, harmonization of the statutory audit function in the EU. A major aim was to enhance a uniform level of high audit quality across the EU member states. In the aftermath of the financial crisis of 2007–2008, the European Commission (EC) questioned the adequacy of its legislative auditing framework in the EU. The EC launched a Green Paper (European Commission, 2010) to open the debate on potential measures to further enhance the audit function in order to contribute to financial stability.[6] The EC stressed the key function of auditing in

6. For discussions of the Green Paper, see Brasel et al. (2011), Humphrey et al. (2011) and Quick (2012).

re-establishing trust and market confidence, and its contribution to financial stability, investor protection, and the reduction of costs of capital. Particular emphasis was given to auditor independence as a core value of statutory auditing, as well as the risks caused by the high supplier concentration in the audit market (Kohler, Quick and Willekens, 2016). After the publication of the Green Paper, the European Commission initiated a public consultation, which finally resulted in the amendment of the Directive on statutory audits of annual accounts (European Union, 2014a) and the EU Regulation on specific requirements regarding statutory audit of public-interest entities (European Union, 2014b). Other key aspects of this EU regulation include the independent oversight (auditor inspections) on the activities of statutory auditors and audit firms, and the launch of extended auditor reporting on the financial statements of public interest entities.

10.3.1 AUDITOR REGULATION AND AUDIT QUALITY

All audit regulations share at least one common objective: they aim to improve and/or safeguard a high level of audit quality. However, regulators do typically not specify what constitutes audit quality. Before discussing various key auditor regulations, we zoom in on the concept of audit quality itself. In the auditing literature, audit quality is often defined as the *ability* of the auditor to detect material misstatements in the financial statements (which depends on the auditor's competence) and his/her *willingness* to issue an appropriate audit report based on the audit findings (which depends on his/her independence). A standard reference is DeAngelo (1981: 186), who defines audit quality as the "market-assessed joint probability that a given auditor both discovers (a) breach in the client's accounting system, and (b) reports the breach." A feature that characterizes audit quality is that it is *unobservable* for parties not involved in the audit, including regulators, as the details of the audit (production) process, such as audit planning, risk assessments, performed audit procedures, and evaluation of audit evidence, are not publicly disclosed (Eilifsen and Willekens 2008). The only observable output of the statutory audit is typically the audit report, and in most cases, this is an unqualified (clean) opinion including boilerplate jargon.

Causholli and Knechel (2012) argue that the audit service exhibits credence features, and hence can be seen as a "credence good", which implies that not only distant shareholders and stakeholders, but even the audited company (management) itself, cannot observe the audit quality supplied. Their arguments are as follows:

Two aspects of the audit production process suggest that the audit may have

> significant credence attributes. First, the outcome of an audit is unobservable. The audit risk model is based on the assumption that the residual risk that the auditor will fail to detect one or more material misstatements always exists (AICPA1983) so the actual level of assurance achieved can never be known (O'Keefe et al., 1994; Knechel et al., 2009). Second, the idiosyncratic and uncertain nature of the audit process means that only the auditor can decide how much effort to exert and evidence to gather to satisfy professional auditing standards. The auditor diagnoses the extent of service required (planning) and provides the actual service (testing). Although, the auditee may have some insight into his/her own risk of material misstatements, the auditor establishes the audit scope based on professional judgment.

DeFond and Zhang (2014) also emphasize that audit quality is difficult to measure because the amount of assurance auditors provide is unobservable. On the contrary, various consequences and characteristics of the audit process are observable. DeFond and Zhang (2014) argue that audit quality improves financial reporting quality by increasing the credibility of the financial reports. As a result, audit quality is a component of financial reporting quality, and it is difficult to distinguish between the two. It is important in this context that financial reporting quality is not only determined by audit quality, but also—and even mainly—by the firm's financial reporting system and the firm's innate characteristics, such as the quality of its operations and governance. DeFond and Zhang (2014) make a taxonomy of audit quality proxies used in the audit literature and distinguish between two categories: 1) measures based on the output of the audit process, such as auditors' reporting conservatism (e.g. going concern reporting for distressed firms), and financial reporting quality; and 2) input-based measures, such as auditor type or audit fees.

All in all, it is widely believed that audit quality is a multi-faceted concept that is largely unobservable. As a result, we argue that it is largely unobservable whether (stricter) auditor regulations actually enhance high audit quality (or some aspects of it), even though regulators claim they do.

10.3.2 AUDITOR INDEPENDENCE REGULATIONS

To enhance high audit quality, regulators typically impose auditor independence requirements. Two important such requirements are: 1) prohibition of the joint supply of certain types of non-audit services to audit clients; 2) mandatory rotation of the lead audit partner, as well as the audit firm, at specified intervals.

Prohibition of the supply of non-audit services. The joint supply of audit and non-audit services by the (incumbent) auditor has been a topic of debate for many years. The 2014 European Regulation tightened the prohibition of the provision of non-audit services by auditors for PIEs. A blacklist of prohibited non-audit services[7] was introduced, as well as the pre-approval requirement by the audit committee for the provision of other non-audit services. Furthermore, a cap is placed on the fees auditors are allowed to earn related to non-audit services: these can maximally amount to 70% of the average audit fees earned on the audit engagement during the previous three years. Note that the Sarbanes-Oxley Act in the United States also bars auditors from providing non-audit services, but that no cap on fees from non-audit services has been introduced.

Mandatory audit partner and audit firm rotation. The 2006 EU Statutory Audit Directive prescribed that lead partners on audit engagements in public interest entities be rotated at least every seven years. Note that in the United States, a *partner* rotation rule of five years has been in place since the Sarbanes-Oxley Act. According to critics, partner rotation may not be a sufficient means to enhance auditor independence. Accordingly, the 2014 EU Regulation further tightened rotation rules for PIEs in the EU as it prescribes that *audit firms* should be rotated at least every 10 years. By way of derogation, Member States may extend audit firm tenure to 20 years where a public tendering process for the statutory audit is conducted, or to 24 years in the case of joint audits. While mandatory audit firm rotation was mainly installed to enhance auditor independence, it could actually also reduce auditor concentration if it provides opportunities to mid-tier and smaller firms to compete with the Big Four. Note that there is no audit firm rotation requirement in the United States. Even though the average audit firm tenure in publicly listed firms is not higher than the 10- to 20-year limits imposed by the EU, some firms stay with the same audit firm for a very long period of time.[8] One notable example is Barclays, which had PwC as its auditor for a period of 120 years. In 2015, PwC was replaced by KPMG in order to comply with the new EU regulation (Wallace 2015). For banks and other large or complex companies, long audit tenure is a rational decision as there are significant switching costs involved when clients hire a new auditor.

7. Examples of services on the blacklist: Tax services; designing/ implementing internal control systems related to financial information; valuation; and services linked to financing and capital structure (for further details, see e.g. Ratzinger-Sakel and Schönberger 2015).
8. The average audit firm tenure is 7.3 years (5.4 years) in countries classified as high (low) litigation risk countries in Brooks, Cheng, Johnston, and Reichelt (2017).

10.3.3 REGULATORS' CONCERNS ABOUT THE AUDIT MARKET STRUCTURE AND LACK OF COMPETITION AMONG AUDIT SUPPLIERS

Regulators around the world have repeatedly expressed concerns about the high level of supplier concentration in the audit market, and question whether the degree of competition is sufficient. Article 27 of the 2014 EU Regulation (European Union, 2014b) addresses the monitoring of market quality and competition, and prescribes the European Competition Network (ECN) to regularly monitor the developments in the market, and in particular to monitor market concentration levels (amongst other things). Article 17 of the Regulation also included the requirement of mandatory audit firm rotation. In the USA, on the contrary, the GAO report (2008) was much milder as there were no recommendations formulated to address auditor market concentration: "The level of market concentration also does not appear to be affecting audit quality as many of our survey respondents and those we interviewed said that audit quality had improved, which some attributed to the Sarbanes-Oxley Act" (GAO report, 2008, p. 5).

The ongoing concerns about supplier concentration in the audit market are a result of the consolidation in the audit industry, which mainly took place during the last two decades of the last century. The last big incident in this context was the collapse of Andersen in 2002. While supplier concentration in the audit market is definitely high, it should not be confused with a lack of competition in that market. From the industrial organization literature we indeed know that Cournot competition models show that market concentration could proxy for competition. Cournot models, however, assume that products are homogeneous (Cabral, 2000) and that suppliers compete on quantity (that is, suppliers are price takers). When suppliers compete on prices, Bertrand models of oligopoly are more appropriate. Consistent with this view, Dedman and Lennox (2009) and Numan and Willekens (2012) argue that there are both theoretical and empirical problems with using concentration as a measure of competition. Note that there is some evidence of (imperfect) oligopolistic competition in the audit market. For example, Numan and Willekens (2012) provide empirical evidence that Big Four audit firms compete by product differentiation as they are able to charge higher fees when they are industry experts in the client's industry, but in addition to that, they are also able to charge an additional fee premium the larger the market power they have vis-à-vis their closest competitor.

10.3.4 INDEPENDENT OVERSIGHT ON THE ACTIVITIES OF STATUTORY AUDITORS AND AUDIT FIRMS

In this section, we zoom in on the question of what mechanisms are in place to ensure that audit firms fulfill their duties and perform independent audits in accordance with applicable regulations, as the new regulations have also significantly altered the system of auditor surveillance. Prior to the Enron scandal in the U.S., the auditing profession used their own system of peer review to ensure that members of the profession adhered to the professional standards. In many countries in the world, similar national systems of peer review existed. This era of self-regulation ended in the U.S. after the accounting scandals of Enron, WorldCom and Tyco, and the subsequent introduction of Sarbanes-Oxley Act in 2002. Paramount in this context was the establishment of the Public Company Accounting Oversight Board (PCAOB), which is one the first independent audit regulators in the world set up to "*oversee the audits of public companies in order to protect the interests of investors and further the public interest in the preparation of informative, accurate, and independent audit reports*".[9] The PCAOB inspects audit firms that have more than 100 listed clients annually, while audit firms with less than 100 listed clients are inspected triennially. In the U.S., there is no independent oversight of audit firms that only audit non-listed firms.

Other countries followed the example of the U.S. with regard to independent oversight over the audit profession, and changed their view about the appropriateness of self-regulation of the audit profession. As a result, oversight bodies that are independent from the national audit profession have been established worldwide since 2000. In particular, 52 independent audit regulators are today members of the International Forum of Independent Audit Regulators (IFIAR), which was established in 2006. The mission of IFIAR "... is to serve the public interest and to enhance investor protection by improving audit quality globally. The overall objective is to: 1) Share knowledge of the evolving audit environment and the practical experience of independent audit regulatory activity. 2) Promote collaboration and consistency in regulatory activity. 3) Provide a platform for dialogue with other international organizations interested in audit quality."[10]

It is interesting to note that the responsibility of auditor oversight and surveillance remained a national matter within the EU. This is somehow inconsistent with several other regulations that moved up to the European level, such as the mandatory application of "International Financial Reporting Standards (IFRS) as

9. https://pcaobus.org/About/History/Pages/default.aspx
10. https://www.ifiar.org/about/#who-we-are

approved by EU" for PIEs as a common reporting language in the financial statements, without any possibility for EU Member States to insert individual adjustments. As for auditor oversight, however, the new 2014 regulation only mandates that each country has a "competent body" with responsibility for oversight of auditors. The EU has also established the Committee of European Auditor Oversight Board (CEOAB). The role of CEOAB "is to strengthen EU-wide audit oversight, which is a key objective of the new EU legislation on statutory audit that took effect on 17 June 2016."[11]

10.3.5 EXTENDED AUDITOR REPORTING REQUIREMENTS FOR PUBLIC INTEREST ENTITIES

Historically, auditor reports have used standard language merely attesting to whether the audited financial statements are consistent with the relevant GAAP (Generally Accepted Accounting Principles) or not, and no information is provided about the potential risks an audited entity may face (except for the going concern risk). Over the past decade, various regulators and standard setters, such as the European Commission, the International Auditing and Assurance Standards Board (IAASB) and the PCAOB, have started initiatives to improve the auditor's reporting model and to enhance transparency. In the EU, new regulation on extended auditor reporting has applied since 17 June 2016, i.e. for audits of financial statements for periods beginning on or after that date. Concretely, the enhancement of the auditor's report draws on developments at the international audit standard setting level and the new Auditor Reporting Model (particularly ISA 700 Revised and ISA 701) issued by the International Auditing and Assurance Standards Board (IAASB) and which, in the meantime, has also been introduced by the PCAOB in the US. The main feature of the new Auditor Reporting Model is the introduction of so-called Key Audit Matters (KAM). The identification and communication of KAM according to ISA 701 is mandatory for the audits of listed entities only. A description of the Key Audit Matters includes the most significant assessed risks of material misstatement, including assessed risks of material misstatement due to fraud, a summary of the auditor's response to those risks, and, where relevant, key observations arising with respect to those risks and reference to the relevant disclosures in the financial statements (European Union, 2014b). Further enhancements (beyond the Key Audit Matters) to the auditor

11. https://ec.europa.eu/info/business-economy-euro/banking-and-finance/financial-reforms-and-their-progress/regulatory-process-financial-services/expert-groups-comitology-and-other-committees/committee-european-auditing-oversight-bodies_en

report include a statement on auditor independence and an explanation of the extent to which the statutory audit was considered capable of detecting irregularities, including fraud. Even if the EU regulation does not explicitly use the notion of KAM, it is widely recognized that the EC requirements and the approach taken by the IAASB are generally consistent.

10.4 FUTURE RESEARCH OPPORTUNITIES: IDEAS AND CHALLENGES

The audit profession defines the purpose of an audit to "enhance the degree of confidence of intended users in the financial statements" and to express an opinion "on whether the financial statements ... give a true and fair view in accordance with the [accounting] framework" being used to prepare the financial statements.[12] To help ensure that this purpose is met, the accounting profession, as well as policy makers and regulators, have implemented many standards and rules that govern the audit process. However, it is not at all clear whether the regulations work as intended, and whether the recent more burdensome and costly regulations have indeed improved audit quality. In this section, we very briefly point to some critical issues for research, as well as challenges that hinder such research.

Empirical cause and effect analyses of audit regulations. The recent changes in the EU regulation enable researchers to observe and study how and whether auditor behavior changed in the aftermath of these new regulations. Regulatory changes make it possible for researchers to capture the relationship between causes (for example, lack of auditor independence) and consequences (audit quality) in a cleaner way. Relevant research questions are the following: Does mandatory audit firm rotation indeed increase audit quality? Does a cap on fees earned from non-audit services by the incumbent auditor increase audit quality? Does audit quality improve after a country has strengthened its system for surveillance of auditors, and is the improvement related to the strength of a particular legal system and/or other characteristics specific to the country, company, or audit firm? Does the new extended format of the audit opinion increase investors' confidence in companies' financial statements? As time goes by, and as more observations from the post-implementation period accumulate, researchers will be able to assess whether the aims of the new regulations are being met.

Looking beyond first order effects of regulatory changes via analytical research. Policy makers and regulatory bodies seem to have a tendency to prioritize first order effects—that is, the immediate effects we observe after new regu-

12. http://www.ifac.org/system/files/downloads/a008-2010-iaasb-handbook-isa-200.pdf

lations are installed. However, immediate effects may cause other things to change as well, or, in other words, second order effects are likely too. Analytical research has the potential to analyze both types of effects the regulatory changes may bring about. An example of such analytical research is a study by Bleibtreu and Stefani (2017) on mandatory audit firm rotation. In the U.S., the United Kingdom, and many EU countries, more than 90 percent of large listed firms are audited by the Big Four audit firms (Francis et al., 2010), and regulators are concerned about the high level of supplier concentration in this market. One concern is that the high market concentration of Big Four audit firms represents a systemic risk in the audit market and hence a threat to financial stability. At the same time, regulators are concerned that auditors are becoming too familiar with their clients and thus not able to perform independent audits. As a means to both decrease audit market concentration and increase auditor independence, the EU introduced mandatory audit firm rotation. Bleibrue and Stefani (2017), however, show that the regulators' goals of simultaneously decreasing client importance and audit market concentration are in direct conflict. Thus, the sum of first and second order effects may be different from what was expected by the proponents. Analytical modeling may clarify under which conditions regulations can be expected to have the desired outcome, and when there are reasons to believe that adverse effects may dominate.

Secrecy, a major hindrance to auditing research's full potential. A key challenge in auditing research is the lack of access to relevant data from regulatory bodies and audit firms. For example, without access to working papers from audit firms, it is not possible—or at least, it is very difficult—to assess the effect of the audits on, for instance, earnings quality and earnings management (Lennox et. al 2015), or to understand why engagement partners within the same audit firms deliver audits of different quality (Gul et. al. 2013).

Regulators in different countries inspect audit firms and a selection of their clients on a regular basis. Through the inspections, the regulatory bodies gather knowledge about audit firm behavior and discover areas with deficiencies. Their work and findings are surrounded by secrecy and confidentiality.[13] Researchers and regulators (national bodies as well as IFIAR and CEOAB) could work

13. This causes problems not only for researchers, but also for regulators themselves, because audit quality inspections may involve audit firms located outside the jurisdiction of the regulatory body. "Many American companies have major operations in China, and what goes on in those audits is quite opaque to US regulators and investors" according to Lewis H. Ferguson (Chair of the Global Public Policy Committee Working Group of the International Forum of Independent Audit Regulators (IFIAR) (Tapestry Networks, 2015, p. 6).

together to analyze the effectiveness of inspections. Are some inspections more efficient than others, and why? To whom should findings be communicated, and in what form? Would public disclosure of audit deficiencies strengthen audit firms' incentives to provide audits of sufficiently high quality? There are substantial variations among countries in how inspections are conducted and how findings are communicated, and lessons will be learned only if researchers are able to tap into the rich knowledge base that resides with the regulators and the audit firms.

10.5 CONCLUSION

The shock created by the Enron scandal spurred increased auditor regulation and surveillance of listed firms and their auditors. The financial crises in 2007 and 2008 once again renewed regulators attention on auditors, as the latter "... gave no warning of the banking crises" (UK House of Lords, 2011: 5). In an inquiry held by the House of Lords in the United Kingdom, a representative of the Big Four audit firms stated that they had carried out their duties properly. The House of Lords commented (p 40): "In the light of what we now know, that defence appears disconcertingly complacent. It may be that the Big Four carried out their duties properly in the strictly legal sense, but we have to conclude that, in the wider sense, they did not do so."

The dissatisfaction of policy makers with how the market for auditing services operates has led to new and tighter auditing regulation. In this chapter, we have described the most important initiatives. Researchers have the opportunity to analyze to what extent the new regulation increases audit quality, but in order to do so they have to overcome a serious hurdle: Access to relevant auditing data. As we argued above, audit quality is for the most part unobservable, except for those that actually do the audit and those that inspect the auditors. Thus, in order to gain a better understanding of how audit quality is impacted by regulation, we call on audit firms, regulators, and researchers to join forces, collaborate and share experiences, insight, knowledge and data.

REFERENCES

Bleibtreu, C. and U. Stefani. 2018. The Effects of Mandatory Audit Firm Rotation on Client Importance and Audit Industry Concentration. *The Accounting Review*, (93)1, pp. 1–27.

Brasel, K. R., Doxey, M. M., Grenier, J. H. & Reffett, A. 2016. Risk disclosure preceding negative outcomes: the effects of reporting critical audit matters on judgments of auditor liability. *The Accounting Review*, 91(5), pp. 1345–62.

Brooks, L., A.C.S. Cheng, J. Johnston, and K. Reichelt. 2017. Estimates of Optimal Audit Firm Tenure Across Different Legal Regimes. *Journal of Accounting, Auditing & Finance.* 32(1) pp. 3–39.

Cabral, L.M.B. 2000. R&D cooperation and product market competition. *International Journal of Industrial Organization*. 18(7), pp. 1033–1047.

Causholli and Knechel (2012) An examination of the credence attributes of an audit. *Accounting Horizons*, 26(4), pp. 631–656.

Cooter, R. D. and Ulen, T. (1988), Law and Economics (Glenview, Illinois: Scott, Foresman).

Dedman and Lennox (2009. Perceived competition, profitability and the withholding of information about sales and the cost of sales. *Journal of Accounting and Economics*. 48(2-3), pp. 210–230.

DeFond, M. and J. Zhang. 2014. A review of archival auditing research, *Journal of Accounting and Economics* 58(2-3), pp. 275–326.

DeAngelo, L. E. 1981. Auditor Size and Audit Quality. *Journal of Accounting and Economics*, 3(3), pp. 183–199.

Eilifsen A. and M. Willekens. 2008. In the name of trust: some thoughts about trust, audit quality and audit regulation in Europe. In: *Auditing, trust and governance: developing regulation in Europe*, (pp. 1–18). U.K.: Routledge.

European Commission. 2010. *Green Paper on Audit Policy: Lessons from the Crisis*. Brussels: European Commission.

European Union. 2006. *Directive 2006/43/EC of the European Parliament and of the Council of 17 May 2006 on statutory audits of annual accounts and consolidated accounts, amending Council Directives 78/660/EEC and 83/349/EEC and repealing Council Directive 84/253/EEC*. Brussels: European Union.

European Union. 2014a. *Directive 2014/56/EU of the European Parliament and of the Council of 16 April 2014 Amending Directive 2006/43/EC on Statutory audits of Annual Accounts and Consolidated Accounts*. Brussels: European Union

European Union. 2014b. *Regulation (EU) No 537/2014 of the European Parliament and of the Council of 16 April 2014 on Specific Requirements Regarding Statutory Audit of Public-Interest Entities*. Brussels: European Union.

Francis, J.R. 2004. What do we know about audit quality? *The British accounting review* 36 (4), 345–368.

Francis, J., P. Michas, and S. Seavey. 2010. *Big Four Audit Market Competition and Earnings Quality around the World*. Working paper, University of Missouri.

GAO. 2008. *Audits of Public Companies: Continued Concentration in Audit Market for Large Public Companies Does Not Call for Immediate Action. Report to Congressional Addresses*. United States Government Accountability Office, Report Number GAO-08-163.

Gul, F.A., D. Wu, and Z. Yang. 2013. Do Individual Auditors Affect Audit Quality? Evidence from Archival Data. *The Accounting Review*, 88(6) pp. 1993–2023.

Hay, D. R.W. Knechel and M. Willekens. 2014. Introduction: The function of auditing. Ch. 1 in Hay, D. R.W. Knechel and M. Willekens (eds.), *Routledge Companion to Auditing*. Routledge. Addington and New York. p. 1—10.

UK House of Lords. 2011. *Auditors: Market Concentration and their Role.* London: Authority of the House of Lords.

Humphrey, C., Kausar, A. and Loft, A. 2011. Regulating audit beyond the crisis: a critical discussion of the EU Green Paper. *European Accounting Review*, 20(3) pp. 431–57.

ISA 700 (n.d). International Standard on Auditing 700 (Revised): *Forming an Opinion and Reporting on Financial Statements.* The International Federation of Accountants.

ISA 701 (n.d). International Standard on Auditing 701 (New): *Communicating Key Audit Matters in the Independent Auditor's Report.* The International Federation of Accountants.

Just, R. E., Hueth, D. L. and Schmitz, A. 1982. *Applied Welfare Economics and Public Policy*. Englewood Cliffs, NJ: Prentice Hall, p. 491.

Khurana, I.K., and K.K. Raman. 2004. Litigation risk and the financial reporting credibility of Big 4 versus non-Big 4 audits: evidence from Anglo-American countries. *The Accounting Review*. 79(2), pp. 473–495.

Kohler, A.G., R. Quick and M. Willekens 2016. The New European Audit Regulation Arena: Discussion of New Rules and Ideas for Future Research. *International Journal of Auditing*, 20, pp. 211–214.

Langli, J.C. and Svanström, T., 2014. Audits of private companies. Ch. 12 in: D. Hay, W.R. Knechel, and M. Willekens, eds. *Routledge Companion to Auditing.* Routledge, Addington and New York. pp. 148–159.

Lennox 1999; Audit Quality and Auditor Size: An Evaluation of Reputation and Deep Pockets Hypotheses. *Journal of Business Finance & Accounting*, 26(7), pp. 779–805.

Lennox, C., Z. Wu, and T. Zhang. 2015. The effect of audit adjustments on earnings quality: Evidence from China, *Journal of Accounting and Economics*, 61(2-3) pp. 545–562.

Lennox, C. and X. Wu. 2017. A Review of the Archival Literature on Audit Partners. *Accounting Horizons*, forthcoming.

Narayanan, V.G. 1994. An Analysis of Auditor Liability Rules. *Journal of Accounting Research*, 32, Studies on Accounting, Financial Disclosures, and the Law, pp. 39–59.

Numan, W., and M. Willekens. 2012. An empirical test of spatial competition in the audit market. *Journal of Accounting and Economics,* 53: 450–465.

Ogus, A. I. 1994. *Regulation. Legal Form and Economic Theory*. Oxford, Clarendon Press.

Quick. R. 2012. EC Green Paper proposals and audit quality, *Accounting in Europe*, Vol. 9, No. 1, pp. 17–38.

Ratzinger-Sakel, N.V.S. and M.W. Schönberger. 2015. Restricting Non-Audit Services in Europe—The Potential (Lack of) Impact of a Blacklist and a Fee Cap on Auditor Independence and Audit Quality. *Accounting in Europe*, 12(1) pp. 61–86.

Schwartz, R. 1998. Auditors' Liability, Vague Due Care, and Auditing Standards. *Review of Quantitative Finance and Accounting*, 11(2), pp. 183–207

Tapestry Networks, 2015, The evolution of audit regulators, ViewPoints for the Audit Committee Leadership Summit, July 13. http://www.ey.com/Publication/vwLUAssets/EY-the-evolution-of-audit-regulators/$FILE/EY-the-evolution-of-audit-regulators.pdf

Shavell, S. 1984. Liability for Harm Versus Regulation of Safety, *Journal of Legal Studies*, 13, pp. 357–74.

Ulen, T.S. 1993. *Cognitive Imperfections and the Efficiency of Legal Rules*. Working Paper, University of Illinois.

Vanstraelen, A. and C. Schelleman. 2017. Auditing private companies: what do we know? *Accounting and Business Research*, 47(5) p. 565–584.

Wallace, T. 2015. Barclays hires KPMG as its auditor after 120 years with PwC. https://www.telegraph.co.uk/finance/newsbysector/epic/barc/11716302/Barclays-hires-KPMG-as-its-auditor-after-120-years-with-PwC.html. July 3.

Watts, R.L. and J.L. Zimmerman. 1986. *Positive Accounting Theory.* Englewood Cliffs, N.J. Prentice-hall.

Willekens, M., A. Steele and D. Miltz. 1996. Audit Standards and Auditor Liability: A Theoretical Model. *Accounting and business research*. 26(3), p. 249–264.

Part III

Deciding, incentivizing and innovating

DOI: 10.18261/9788215031583-2018-12

Chapter 11
From Creativity to Innovation:
Four Leadership Lessons about Capitalizing on High-Potential Ideas

MIHA ŠKERLAVAJ

ABSTRACT Creative ideas fuel modern organizations and are increasingly salient in times of change. However, novelty—one defining characteristic of creative ideas—is associated with risk. That being said, highly creative ideas tend to represent the most potential, relative to the value they add to organizations and their members. How can leaders increase the odds of successfully transforming high-potential creative ideas into innovative realities? This chapter reviews the most current research findings on optimizing high-potential creative ideas to render the innovation advances they promise. It summarizes and exemplifies the following four leadership lessons: 1) change agents, 2) supportive leadership, 3) integrating multiple perspectives from assorted stakeholders, and 4) facilitating creative employee behavior in the workplace. Research suggests that effectively capitalizing on high-potential ideas in organizational settings requires active leadership that involves a mastery of the competencies of relevant change agents, as the development of new ideas requires rigorous in-context management of the change process. Leaders need to show two-dimensional support of tasks and individuals, not only to provide resources and assistance as needed, but also to facilitate proactive behaviors by challenging employees to depart from the status quo. The successful leader, above all, recognizes that capitalizing on creativity is a social process that requires contributions from multiple viewpoints, and that various stakeholders need to be involved.

KEYWORDS: leadership | creativity | innovation | change agent

11.1 INTRODUCTION

> We can't solve problems by using the same kind of thinking we used when we created them.
>
> – *Albert Einstein*

The BI Norwegian Business School's celebration of its 75 years of existence corresponds with many unprecedented challenges that we face as societies, organizations, and individuals. Some literature (Rittel and Webber, 1973) refers to issues as "wicked problems", a term that refers to a class of ill-formulated social-system problems characterized by ambiguous information, multiple clients and decision makers with conflicting preferences, and thoroughly confusing ramifications for and within the whole system. At the societal level, the United Nations (UN, 2015) declared 17 sustainable development goals (accompanied by 149 targets) that, with a vast scope, including the following, as expected: eradication of poverty, zero-level hunger, universal good health and well-being, clean energy and water, protective climate action, reduction of inequalities, decent work opportunities. However, the UN's list of goals included additional items that represent complicated and higher-order goals, such as promoting sustainable industrialization and fostering innovation, to name just few. At an organizational level, strategic priorities include digital transformations, technological advances (artificial intelligence and big data), the automation and robotization of work, disruptive business models, gradually-evolving organizational cultures, new agile and design-oriented methods of working and thinking, and the development of change-agency capacity. Individuals perceive such goals as dual challenges, reflecting both opportunities and threats that impose a need to make sense and meaning of new realities. The cards are being reshuffled, as some authors assert that we now face the fourth S-curve or a fourth industrial revolution, powered by new raw material (data), new machines (systems of intelligence, particularly artificial ones), and new business models (Frank, Roehrig, and Pring, 2017).

Although the UN explicitly frames innovation as one of its sustainable development goals, I argue that creativity and innovation are as integral to the problem as the solution. Much of the trouble with innovation processes derives from the so-called innovation paradox, which refers to the following observation: while innovation requires both the creation and implementation of novel ideas, the same conditions that favor the creation of novel ideas often impede the process of implementing those ideas (Miron-Spektor, Erez, and Naveh, 2011). While new societal, organizational, and individual challenges require innovative solutions, creativity has yet to become a sufficient precondition for the resolution of our societal challenges.

Despite their potential utility, high-potential ideas aimed at solving wicked problems typically depart from the status quo, and therefore tend to be perceived as radically novel, even threatening. They square with many intuitive organizational and psychological barriers, including the "not-invented-here syndrome" (Katz and Allen, 1982) and middle status conformity (Phillips and Zuckerman, 2001). Consequently, high-potential ideas are less likely to be deployed than ideas that are only moderately novel—and thereby represent lower potential for innovative solutions (Škerlavaj, Černe, and Dysvik, 2014). One particularly vivid example is entomophagy, the emerging industry of edible insects (Nurmohamed and Harrison, 2016), which has the potential to feed many more individuals in a much more environmentally-sustainable way than our current nutritional strategies. While the practice of eating insects may be less unusual in Eastern cultures, Western cultural reactions to the prospect tend to be strongly negative, often involving a sense of revulsion. To wit, while we certainly need innovative ideas, the simple novelty of such ideas will not suffice.

From a perspective that regards organizations as platforms for coordinated human activity (Barnard, 1968), I invite the reader to focus on the role organizational leaders assume in the process of capitalizing on high-potential ideas. The central research question of this chapter revolves around how organizational leaders can promote successful capitalization on highly creative ideas with the potential to resolve the challenges of the modern world. This cutting-edge literature review is based on a book project (Škerlavaj, Dysvik, Černe, and Carlsen, 2016) generated by an impressive spectrum of thinkers—42 researchers from 28 organizations, in 13 countries across four continents—all of whom focus on identifying the actions most capable of effectively deploying the creativity intrinsic to high-potential ideas that depart significantly from the status quo. This chapter narrows its focus to interpret four critical lessons for innovative leadership, summarize current literature about the most advanced modes of employing the innovation process, and furnish implications for practical leadership strategies and techniques.

11.2 LEADING HIGH-POTENTIAL CREATIVE IDEAS TO INNOVATION

Capitalizing on creativity is not analogous to traveling down a one-way street embellished with sequential moments of glory, where ideas can grow from birth to realization, without regard for disruptions or obstacles (Černe, Carlsen, Škerlavaj, and Dysvik, 2016). Actual journeys from creativity to innovation (Van de Ven, 1999) are far messier, because ideas are seldom static when they confront and

interact with the people interested in applying them. In fact, it is far from unusual to generate metaphors related to creative and innovative work that refer to maze behavior, or hiking toward a hostile, mountainous terrain that keeps the ultimate goal hidden from sight (Fisher, Pillemer, and Amabile, 2017).

Regardless of the domain, most research on creativity regards it as a combination of novelty and potential utility (Amabile, 1996; George, 2007; Simonton, 2004). Intrinsic to the very concept of creativity is the belief that, for an idea to be authentically creative, it must be deployed in the field and highly valued by key stakeholders (Csikszentmihalyi, 1999); mere novelty is not adequate enough. In agreement with recent literature reviews on idea generation and idea implementation (Anderson, Potočnik, and Zhou, 2014), and aiming to theorize about novelty and utility as dimensions of creativity (Berry, 2012), I propose the term *high-potential ideas* to refer to ideas that are, first, highly novel, relative to the mainstream and second, have high potential utility for the key stakeholders. Paradoxically, key stakeholders are often unaware or unwilling to recognize the utility of high-potential creative ideas, as illustrated by the foregoing example of the entomophagy industry. The perceived potential utility and the actual potential utility of ideas are rarely the same. Furthermore, ideas are rarely born in their most novel or useful forms; rather, they become novel or useful, as people work on them—expanding, molding, amplifying, reiterating, sharpening, and intermingling them in order to distill and intensify their most valuable components and implications (Carlsen, Clegg, and Gjersvik, 2012). As such, capitalizing on high-potential ideas requires achieving mastery of the labyrinthine process by which high-potential ideas are led carefully toward their embodiment as innovations, and that (along with other prerequisites and boundary-condition facilitation) involves active leadership.

What kind of leadership (and what sort of following) is conducive to more effective capitalization on high-potential ideas? The four key lessons for leaders of the innovation process (according to multiple contributions to the recently edited book project) are delineated in Figure 11.1. This chapter therefore aims to elaborate on, and exemplify, each of the four leadership lessons. First, introducing innovations and capitalizing on high-potential ideas requires the consistent championing of ideas (Černe, Kaše, and Škerlavaj, 2016) and the development of change-agency skills that involve a deep understanding of context, process, and the nature of the change that will be necessary to capitalize on the new idea. Second, supportive leadership (Buch and Kuvaas, 2016; Černe, Škerlavaj, and Dysvik, 2016) can improve the odds for successfully capitalizing on high-potential ideas, by providing both task- and role-oriented support related to dimensions of

resource allocation, motivational prompting, and role modeling. Third, capitalizing on creativity is a multi-player game (Mørk and Hoholm, 2016), in which the ability to observe the scene from different angles is essential. Leaders must understand and receive instruction in their role as facilitators, employing different idea-work methods, including agile and design-based thinking (Rauth and Nabergoj, 2016). Fourth, leaders must construct working environments that inspire employees to demonstrate and reach for their highest proactive potential (Hudovernik, Škerlavaj, and Černe, 2016).

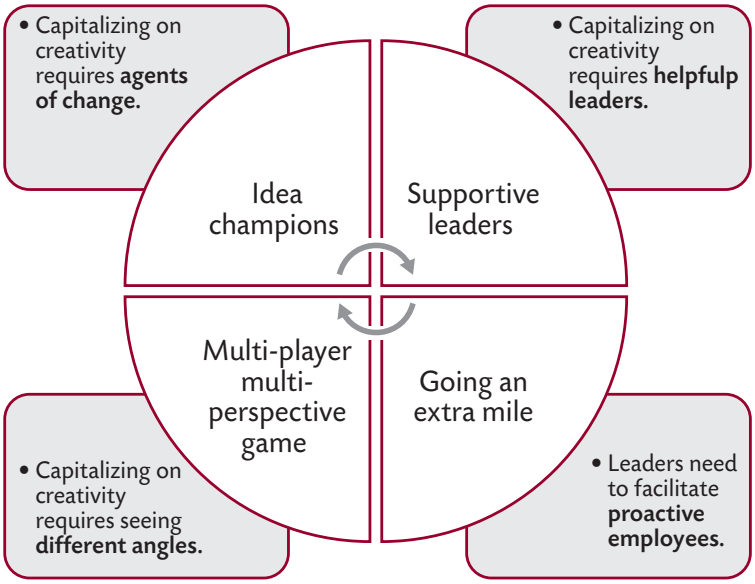

FIGURE 11.1. Four leadership lessons about capitalizing on creativity

11.2.1 LESSON ONE: CAPITALIZING ON HIGH-POTENTIAL IDEAS REQUIRES CHANGE AGENTS

Capitalizing on high-potential ideas is essentially a process of change management. The content of such processes consists of capitalizing on ideas that are perceived as departures from the status quo, and this require leaders to act as agents within the context of this pursuit of change. High-potential ideas are never born in a vacuum, and they invariably compete (and connect) with other ideas for time, man-hours, brainpower, finances, and other resources needed for successful capitalization. High-potential ideas also have consequences for extant ideas and the people associated with them. Schön (1963) goes so far as to claim that ideas must

either find a champion or die. All of this implies the substantial importance that attaches to the interpersonal processes that revolve around successfully capitalizing on a high-potential idea. In practice, human agency represents the realized capacity of people to act upon the world in a purposive and reflective manner, and to acknowledge an emergent need to remake that world, if they will continue to live within it (Inden, 1990).

Markham and colleagues (1991) describe innovation change agents or idea champions as strong advocates for a project who generate positive behavioral support for an innovation during its development in the face of organizational neutrality or opposition. The literature cites a wide variety of forms, names, and special cases of innovation change agents. Internal agents of change stand tall, elevated above the crowd by their vision and focus, and work actively to build coalitions, seek sponsorship, and ultimately aim toward legitimizing an idea. Creative "bootleggers" are those organizational members who take the initiative to work on ideas in the absence of formal support (or even the awareness of upper management) with the aim of benefiting the company (Criscuolo, Salter, and Ter Wal, 2013). Similarly, corporate "smugglers" experiment, evolve their methods, and endeavor to include others, instead of pushing creative ideas through organizations by force (Lempiälä, 2011). Stealth innovators operate under the radar (Miller and Wedell-Wedellsborg, 2013), quietly hidden from supervisors in their pursuit of innovation; this operating mode is exemplified by the change agency involved in the conception of outsourcing tasks, rather than entire jobs, at a global pharmaceutical company.

The change agent arrived at the idea by using his own experience and a human-centric approach toward innovation to unburden knowledge-workers of the grunt work associated with direct on-click outsourcing (e.g., slide preparation, data entry, and analysis) to trusted partners. He stayed under the radar for one year, and actively spent this time developing the proposed service, while accumulating evidence in favor of its utility, and gaining allies to support his ultimate disclosure of the work. Then, he went public, received funds, and even attained a formal role attached to the new service, which he had successfully innovated. While this is a relatively straightforward case, in which a single individual championed the idea he came up with in the first place, one that is analogous to Grant's (2017) idea of originals, many organizational realities involved with capitalizing on high-potential ideas are substantively more complex and often intuitively distinguish between the tasks of idea ownership and idea championing.

Whether above or below the radar, these forms of change agency aimed at capitalizing on high-potential creative ideas entail an extraordinarily high level of

commitment from either leaders or team members. The crux of change and innovation leadership competencies largely consists of timing actions appropriately in a given organizational context in order to determine the best possible fit between actions, context, and content of a high-potential creative idea. The major role of change agents involves the engagement of (Ford, Ford, and D'Amelio, 2008; Sonenshein and Dholakia, 2012) the intended recipients of the change, focused on conquering the shortcomings of still-prevalent top-down approaches to change management (Kotter, 2007), which rarely function effectively when successfully capitalizing on ideas requires winning people's hearts and minds. Luckily, the competencies associated with change agency and change leadership can and should be sharpened by instruction and practice.

11.2.2 LESSON TWO: CAPITALIZING ON HIGH-POTENTIAL IDEAS REQUIRES SUPPORTIVE AND HELPFUL LEADERS

Leaders must serve as change agents for innovation; they also need to empower others to contribute to capitalizing on high-potential ideas. One important way they can do this is by practicing supportive supervision, which can be defined as the degree to which employees believe their supervisors authentically value their contributions and care about their well-being (Eisenberger, Stinglhamber, Vandenberghe, Sucharski, and Rhoades, 2002). It involves both task- and relationship-oriented behavior, in the same way that it involves both instrumental and socio-emotional forms of support (Amabile, Schatzel, Moneta, and Kramer, 2004). Supportive supervisors enable their employees to capitalize on high-potential creative ideas by providing resources, motivational support, and serving as role models.

In terms of resource allocation, supportive leaders provide employees with access to those resources necessary for successful implementation of ideas; providing such resources might involve rendering the optimal amount of time, advice, funds, equipment, and connections to experts and implementers. Motivational support refers to the stimulation of employee perceptions of competence and relatedness, while role modeling increases the perceived challenge (as opposed to hindrance) by having managers demonstrate innovation-related behaviors themselves.

Some recent exemplary behaviors include helping employees navigate complex tasks, (i.e. providing deep help; Fisher et al., 2017), guiding or path-clearing as a way of providing external help with complex tasks related to capitalizing on high-potential ideas, and providing meaningful feedback (Harrison and Dossinger,

2017; Hoever, Zhou, and van Knippenberg, 2017). As one leader puts it "(leading innovation)… is often about removing obstacles (including those in people's minds) and securing resources." (—Marc O'Neill, from Liedtka, 2009).

11.2.3 LESSON THREE: CAPITALIZING ON HIGH-POTENTIAL IDEAS REQUIRES OBSERVING THE SITUATION FROM DIFFERENT ANGLES AND INTEGRATING MULTIPLE PERSPECTIVES

Early work on creativity emphasizes the role of different perspectives in generating creative output. The random variation model of creativity postulates that idea generation (i.e., creativity) derives from divergent inputs that increase variance across ideas, thereby raising the odds that one of the group's ideas will be a radical breakthrough, clearly streaking toward a successful creative product. More recent work (Harvey, 2014) has put forth a theoretical dialectical model that regards the integration of group member perspectives—identified as "creative synthesis"—as the foundation for novel, breakthrough, high-potential ideas. Team members must combine their resources and capacities, at the cognitive (i.e., the abilities of individual members to generate ideas), social (i.e., the specific composition of the group), and environmental (i.e., the extent to which the organizational environment supports and motivates group members) levels to promote the creation of extraordinary output. While creative synthesis is a prerequisite for this output, it does not, by itself, suffice as a mode to ensure capitalization on high-potential ideas. When breakthrough ideas move from creative teams into the broader organizational space, they need help from all available quarters to take flight.

In their study of a global consumer products company, Rauth and Nabergoj (2016) describe *design thinking* (Brown, 2008) as a multi-actor sense-making process for capitalizing on high-potential breakthrough ideas. On this journey, ideas confront many internal and external stakeholders, who often hold conflicting views. One example was a project that dealt with illegally pirated copies of products, aiming to develop ideas about suitable technique to frustrate counterfeiters and cause them to cease their efforts to copy company brands. This organization hosted a series of workshops that derived and applied the following set of practices: 1) iterative engagement of functionally and hierarchically diverse individuals; 2) heterogeneous composition of teams; 3) iterative involvement of individuals, teams, and groups, and 4) exposure to, engagement with, and action within the context of conflicting views (one at the time). As a key leadership lesson aimed at promoting capitalization on high-potential creative ideas, value attaches to design thinking when it is functionally applied as a mode of describing precisely how such capitalization operates when leaders actively facilitate human-centered inno-

vation journeys (Kolko, 2015) aimed at dealing with wicked problems. However, it might be noted here that for a high-potential idea to be worthy of the efforts to capitalize on it, the idea must make sense to a variety of internal and external stakeholders. In turn, this requires handling a set of views that are often conflicting, or at least confused, in an iterative and experimental fashion that resembles several of the innovative working methods suggested by lean startup (Osterwalder and Pigneur, 2010) and the innovator's method (Furr and Dyer, 2014). To ensure that breakthrough ideas actually emerge, it also requires getting stakeholders across organizational hierarchy levels involved at a relatively early point in the process, due to an increased sense of ownership, commitment, and access to cognitive, social, and environmental resources.

Theoretical (Hernaus, 2016) and empirical work (Černe, Hernaus, Dysvik, and Škerlavaj, 2017) related to job characteristics suitable for creative and innovative work, respectively, have explored the differences between job design related to these two disparate types of work. Comparing job characteristics that share, at a minimum, the capacity to facilitate either idea generation or idea implementation reflects a primary distinction between certain types of job design that derives from the disparate social characteristics of certain jobs. Job designs more suitable to (i.e., idea implementation) capitalizing on high-potential ideas require more social interaction, increased task interdependence, and intensified social support; more creatively-focused job designs are comparatively solitary (although conventional notions of the lonely genius no longer actually apply), but recent work on the relational job design model has suggested that creative workers can be motivated by prosocial conceptions of the meaning of their contributions. To wit, relational job designs (Grant, 2007) that promote feedback intrinsic to the job itself, while exposing the creative workers to the beneficiary of such contributions, can serve as powerful sources of motivation that facilitate the implementation-related tasks of capitalizing on creative work for the ultimate purpose of innovation.

Given the international nature of modern work, leaders and employees can benefit from further intercultural collaboration and the development of cultural intelligence (Bogilović, Černe, and Škerlavaj, 2017; Bogilović, Škerlavaj, and Wong, 2016). Cultural intelligence (Earley and Ang, 2003), which refers to the capacity to function effectively in culturally diverse environments, across meta-cognitive, cognitive, motivational, and behavioral dimensions, is a vital personal competency that is crucial to understanding and operationalizing the broad range of perspectives that organizational members must depart from if they are to enjoy success in their collective innovative endeavors. A recent meta-analysis (Taras, Steel, and Kirkman, 2016), which assembled all previous studies in accordance with

Hofstede's typology of national culture, showed that 80% of the variation in work-related cultural values could be observed within countries, leaving only 20% of such variations at the international level; the analysis argues against falsely conflating work culture with national culture. It further implies that no room for cultural stereotyping exists in today's world, and begs for considerably more emphasis on individual sources of cognitive and functional diversity, particularly in the context of work involving the generation and implementation of ideas. Furthermore, diversity does not, by itself, breed creativity (Hoever, Van Knippenberg, Van Ginkel, and Barkema, 2012), let alone make it possible to capitalize on it. Capitalizing on high-potential ideas thereby requires individuals with enough cultural intelligence to thoughtfully examine different perspectives, and leaders capable of making sense and meaning out of them.

11.2.4 LESSON FOUR: LEADERS NEED TO FACILITATE PROACTIVE EMPLOYEE BEHAVIORS

If a multiplicity of perspectives represents the flint of the creative process, than proactive employee behaviors are its fuel. Proactive employee behaviors include such descriptors as self-starting, anticipatory, long-term oriented, and persistent (Frese and Fay, 2001); these characteristics can be illustrated by behaviors like feedback seeking, voice, job crafting, taking charge, issue selling, and building social networks (Lam, Spreitzer, and Fritz, 2014). Proactive behaviors are increasingly regarded as critical components of high-caliber job performance (Crant, 2000), and have been associated with desirable organizational outcomes, including creativity (Ohly and Fritz, 2010). Due to their persistent and resilient nature, proactive employee behaviors are important drivers that contribute to the success of capitalizing on high-potential ideas, as they move through organizational landscapes and beyond.

Hudovernik, Černe, and Škerlavaj (2016) used a grounded theory approach to study three cases from the automotive industry, and managed to develop a model that identified antecedents of the proactive employee behaviors needed to capitalize on high-potential creative ideas. At the individual level, previous research (e.g., Crant, 2000; Parker, Bindl, and Strauss, 2010) showed that individual-level antecedents can be organized within an ability–motivation–opportunity framework. Determinants of proactive employee behavior as organized within this framework are as follows: 1) proactive personality, personal initiative, taking charge (personality/ability); 2) role-breadth, self-efficacy, and cognitive-motivational processes (motivation), as well as 3) flexible role orientation, autonomy,

trust, social ties, task interdependence, task complexity, and coworkers' proactive behaviors (opportunity). However, relying on personal dispositions and recruiting employees with proactive personalities may be necessary, but it is hardly sufficient. As a rule, organizations that are successful in capitalizing on high-potential creative ideas have a set of diverse and specific institutional practices that are aimed at motivating proactive employee behaviors. Some of the better-known examples of these enshrined practices include lean workshops that create a bridge between business functions, 20 keys for workplace improvements, ideation workshops and hackathons, IT platforms for innovation combined with off-line activities (such as innovation month), and daily audits; needless to say, these are a mere few among many. All such practices are coupled with personnel, one of who might be the designated innovation process sponsor, including a pool of trained innovation facilitators. Finally, yet crucially, top management philosophy in support of innovation complements all of the abovementioned antecedents to employee proactivity by communicating, visualizing and embodying innovation as an organizational value. Leaders that rely on employee proactivity should regard our model as a system of interconnected-activities operating on multiple levels: individual, team, and organizational.

11.3 PRACTICAL IMPLICATIONS

Research suggests that at least four groups of leadership competencies are required to successfully capitalize on high-potential creative ideas in organizational settings: active leadership that includes mastery of change agent competencies, as new ideas require contextual change process management; second, supportive leadership in task- and people-oriented dimensions, which not only requires leaders to provide resources and other types of support, but also demands that they challenge employees in a way that promotes movement away from the status quo; third, the facilitation of proactive employee behaviors; and fourth, leadership acknowledgement of the nature of capitalizing on creativity as a social process, which requires the involvement of multiple viewpoints and various stakeholders. Such skills, along with a great many more, comprise the role of leaders who wish to propel high-potential creative ideas on the journey toward rebirth as high-value-added innovative services, products, and solutions.

We can derive several practical implications from this. First, any leader in an organization that depends upon renewal, creativity and innovation, should be trained as a change agent. Vital to excelling at change agency is a nuanced understanding of the psychology of change followers and of change as a social process.

The most reliable approach is neither top-down nor strategy-in-the-clouds in terms of abstraction, but rather engaging and involving. Such training programs, as a rule, use experiential learning methods (i.e., case studies, simulations) in their on-the-job programs for broadened systematic training impact (Dysvik, Carlsen, and Škerlavaj, 2017).

Second, leaders must be selected, continuously trained, developed, and promoted on the basis of their supportive and helping behaviors. Despite the ongoing rhetoric that celebrates supportive leadership, studies (Hogan and Kaiser, 2005) continue to report a lack of such supportive leadership, with as many as two-thirds of respondent employees citing leadership as the worst aspect of their job. Significant change is needed, with greater emphasis on more employee-centered forms of leadership pointing toward the route of highest potential. Among them, one should note servant leadership (Greenleaf, 1977, 2002; Stone, Russell, and Patterson, 2004; Van Dierendonck, 2011) as the single leadership form that is directly focused on its followers. Equally important, however, is transformational leadership, with its focus on organization and strong association with innovative work behaviors (Hammond, Neff, Farr, Schwall, and Zhao, 2011). Concrete examples of behaviors required by leaders to improve the odds of capitalizing on high-potential creative ideas include so-called "deep help" (Fisher et al., 2017), where leaders both support and challenge innovation project members, "guiding a team through a difficult juncture by working with its members in several prolonged, tightly clustered sessions, and/or path-clearing by helping a team address a persistent deficit via briefer, intermittent sessions throughout a project's life" (p. 3).

Third, capitalizing on high-potential creative ideas is a social process and a multi-player game. Leaders need to facilitate collaborative employee behaviors during and around innovation journeys. Some approaches shown to be productive are design thinking (Brown, 2009), relational job design, exposure to primary beneficiaries of work (Grant, 2007), cultural intelligence (Earley and Ang, 2003), reciprocity rings (Baker and Bulkley, 2014) and other tools for collaboration that facilitate perspective taking and heightened empathy (Škerlavaj, 2016). Stepping into the shoes of another person, acknowledging that person's perspective, and determining the most suitable course for the innovation journey are prerequisites to benefiting from cognitive, functional, ethnocentric and other types of diversity, as such practices further increases the odds of capitalizing on high-potential creative ideas.

Finally, yet significantly, innovation journeys require persistence, grit (Duckworth, 2016), resilience, and self-initiated proactive employee behaviors. Selecting people who are predisposed toward proactivity is only part of the story. The

development of organizational practices and top management support that both favor employee proactivity constitutes much of the rest of the story representing the prerequisites for capitalizing on high-potential creative ideas as they meet organizational and environmental boundary conditions.

11.4 FUTURE DIRECTIONS

The journey to discover the ideal role of leadership for capitalizing on high-potential ideas is far from complete. The essential question, then, concerns the best and most necessary direction for the additional research. Some recent examples (e.g., Fisher et al., 2017; Perry-Smith and Mannucci, 2017) make compelling claims for the vast potential outcomes of future research about complex innovative work. First, quantitative and qualitative research traditions will need to collaborate in additional research on capitalizing on high-potential ideas in order to bring practice- and process-based perspectives (Langley, 1999) closer to advances in multilevel theorizing and analysis (Kozlowski, Chao, Grand, Braun, and Kuljanin, 2013). With such a dual, yet integrated, approach, we will be able to understand the process of capitalizing on ideas in the context of a clear conception of how leadership and followership of creative and innovative work actually occur.

Second, when discussing leadership, creativity, and innovation, there should be an enlarged focus on the behaviors of leaders and followers. To what extent are leadership practices significant beyond what they say about intentions, attitudes, or traits? Some of the most recent studies focusing on deep help in complex and knowledge-intensive projects recognize leaders as facilitators. Essential skills, therefore, should involve respectful inquiry (Van Quaquebeke and Felps, 2016), as well as feedback giving and searching (Harrison and Dossinger, 2017; Harrison and Rouse, 2015).

In my opening remarks about the challenges and opportunities our society and organizations currently face, I argued that creativity and innovation could function as both the problem, as well as the solution. Generating breakthrough creative ideas is a job that has barely started. Capitalizing on high-potential creative ideas to address the challenges we all face requires mastering leadership competencies for innovative work. Leaders capable of rendering transformations in universal attitudes toward, for example, eating bugs and the emerging industry of entomophagy, will be leaders capable of addressing and meeting the challenges involved in sustainable development goals.

REFERENCES

Amabile, T. (1996). *Creativity in context*. Boulder, CO: Westview Press.

Amabile, T., Schatzel, E. A., Moneta, G. B., and Kramer, S. J. (2004). Leader behaviors and the work environment for creativity: Perceived leader support. *The Leadership Quarterly, 15*(1), 5–32.

Anderson, N., Potočnik, K., and Zhou, J. (2014). Innovation and Creativity in Organizations: A State-of-the-Science Review, Prospective Commentary, and Guiding Framework. *Journal of management, 40*(5), 1297–1333. doi: 10.1177/0149206314527128

Baker, W. E., and Bulkley, N. (2014). Paying it forward vs. rewarding reputation: Mechanisms of generalized reciprocity. *Organization Science, 25*(5), 1493–1510.

Barnard, C. I. (1968). *The functions of the executive* (Vol. 11): Harvard university press.

Berry, J. (2012). *Do we have creative differences? How we construe creativity influences the salience of novelty and usefulness*.

Bogilović, S., Černe, M., and Škerlavaj, M. (2017). Hiding behind a mask? Cultural intelligence, knowledge hiding, and individual and team creativity. *European Journal of Work and Organizational Psychology*, 1–14.

Bogilović, S., Škerlavaj, M., and Wong, S. I. (2016). Idea implementation and cultural intelligence. In M. Škerlavaj, M. Černe, A. Dysvik, and A. Carlsen (Eds.), *Capitalizing on creativity at work: fostering the implementation of creative ideas in organizations* (pp. 39–50): Edward Elgar Publishing.

Brown, T. (2008). Design Thinking. *Harvard business review, 86*(6), 84–92.

Brown, T. (2009). *Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation.* Harper Business.

Buch, R., and Kuvaas, B. (2016). Economic and social leader–member exchange, and creativity at work. In M. Škerlavaj, M. Černe, A. Carlsen, and A. Dysvik (Eds.), *Capitalizing on Creativity at Work: Fostering the Implementation of Creative Ideas in Organizations*. Cheltenham, UK: Edward Elgar Publishing.

Carlsen, A., Clegg, S., and Gjersvik, R. (2012). *Idea Work*. Oslo: Aschehoug.

Černe, M., Carlsen, A., Škerlavaj, M., and Dysvik, A. (2016). Capitalizing on creativity: on enablers and barriers. In M. Škerlavaj, M. Černe, A. Dysvik, and A. Carlsen (Eds.), *Capitalizing on Creativity at Work: Fostering the Implementation of Creative Ideas in Organizations*. Cheltenham, UK: Edward Elgar Publishing.

Černe, M., Hernaus, T., Dysvik, A., and Škerlavaj, M. (2017). The role of multilevel synergistic interplay among team mastery climate, knowledge hiding, and job characteristics in stimulating innovative work behavior. *Human Resource Management Journal, 27*(2), 281–299. doi: 10.1111/1748-8583.12132

Černe, M., Kaše, R., and Škerlavaj, M. (2016). This idea rocks! Idea championing in teams. In M. Škerlavaj, M. Černe, A. Carlsen, and A. Dysvik (Eds.), *Capitalizing on Creativity at Work: Fostering the Implementation of Creative Ideas in Organizations*. Cheltenham, UK: Edward Elgar Publishing.

Černe, M., Škerlavaj, M., and Dysvik, A. (2016). *Supportive supervision: a crucial factor for unlocking the potential of highly creative ideas perspective.* In. Cheltenham, UK: Edward Elgar.

Crant, J. M. (2000). Proactive behavior in organizations. *Journal of management, 26*(3), 435–462.

Criscuolo, P., Salter, A., and Ter Wal, A. L. (2013). Going underground: Bootlegging and individual innovative performance. *Organization Science, 25*(1287–1305).

Duckworth, A. (2016). *Grit: The power of passion and perseverance*. Simon and Schuster.

Dysvik, A., Carlsen, A., and Škerlavaj, M. (2017). Rings of fire : training for systems thinking and broadened impact. In K. G. Brown (Ed.), *The Cambridge Handbook of Workplace Training and Employee Development* (pp. 471–494). Cambridge University Press.

Earley, P. C., and Ang, S. (2003). *Cultural intelligence: Individual interactions across cultures*. Stanford University Press.

Eisenberger, R., Stinglhamber, F., Vandenberghe, C., Sucharski, I. L., and Rhoades, L. (2002). Perceived supervisor support: Contributions to perceived organizational support and employee retention. *Journal of applied psychology, 87*(3), 565.

Fisher, C. M., Pillemer, J., and Amabile, T. M. (2017). Deep help in complex project work: Guiding and path-clearing across difficult terrain. *Academy of Management Journal*. doi:10.5465/amj.2016.0207

Ford, J. D., Ford, L. W., and D'Amelio, A. (2008). Resistance to change: The rest of the story. *Academy of Management Review, 33*(2), 362–377.

Frank, M., Roehrig, P., and Pring, B. (2017). *What To Do When Machines Do Everything: How to Get Ahead in a World of AI, Algorithms, Bots, and Big Data*. John Wiley and Sons.

Frese, M., and Fay, D. (2001). 4. Personal initiative: An active performance concept for work in the 21st century. *Research in Organizational Behavior, 23*, 133–187.

Furr, N. R., and Dyer, J. (2014). *The Innovator's Method: Bringing the Lean Startup Into Your Organization*. Harvard Business Press.

George, J. M. (2007). Creativity in organizations. In J. Walsh and A. V. Brief (Eds.), *Academy of Management Annals* (Vol. 1, pp. 439–477). New York, NY: Erlbaum.

Grant, A. (2007). Relational job design and the motivation to make a prosocial difference. *Academy of Management Review, 32*(2), 393–417.

Grant, A. (2017). *Originals: How non-conformists move the world*. Penguin.

Greenleaf, R. K. (1977). *Servant leadership*. In: New York: Paulist Press.

Greenleaf, R. K. (2002). *Servant leadership: A journey into the nature of legitimate power and greatness.* Paulist Press.

Hammond, M. M., Neff, N. L., Farr, J. L., Schwall, A. R., and Zhao, X. (2011). Predictors of individual-level innovation at work: A meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts, 5*(1), 90–105.

Harrison, S., and Dossinger, K. (2017). Pliable guidance: A Multilevel Model of Curiosity, Feedback Seeking, and Feedback Giving in Creative Work. *Academy of Management Journal*. doi: 10.5465/amj.2015.0247

Harrison, S., and Rouse, E. (2015). An inductive study of feedback interactions over the course of creative projects. *Academy of Management Journal, 58*(2), 375–404.

Harvey, S. (2014). Creative Synthesis: Exploring the Process of Extraordinary Group Creativity. *Academy of Management Review, 39*(3), 324–343. doi: 10.5465/amr.2012.0224

Hernaus, T. (2016). Job design at the crossroads: From "creative" jobs to "innovative" jobs. In M. Škerlavaj, M. Černe, A. Dysvik, and A. Carlsen (Eds.), *Capitalizing on creativity at work: fostering the implementation of creative ideas in organizations* (p. 17): Edward Elgar Publishing.

Hoever, I. J., Van Knippenberg, D., Van Ginkel, W. P., and Barkema, H. G. (2012). Fostering team creativity: perspective taking as key to unlocking diversity's potential. *Journal of applied psychology, 97*(5), 982.

Hoever, I. J., Zhou, J., and van Knippenberg, D. (2017). Different strokes for different teams: The contingent effects of positive and negative feedback on the creativity of informationally homogeneous and diverse teams. *Academy of Management Journal*, amj. 2016.0642.

Hogan, R., and Kaiser, R. B. (2005). What we know about leadership. *Review of general psychology, 9*(2), 169.

Hudovernik, J., Škerlavaj, M., and Černe, M. (2016). Proactive employee behaviors and idea implementation: three automotive industry cases. In M. Škerlavaj, M. Černe, A. Carlsen, and A. Dysvik (Eds.), *Capitalizing on Creativity at Work: Fostering the Implementation of Creative Ideas in Organizations*. Cheltenham, UK: Edward Elgar Publishing.

Inden, R. B. (1990). *Imagining india*. Indiana University Press.

Katz, R., and Allen, T. J. (1982). Investigating the Not Invented Here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R and D Project Groups. *Rand D Management, 12*(1), 7–20.

Kolko, J. (2015). Design thinking comes of age. *Harvard Business Review, 93*(9), 66–71.

Kotter, J. R. (2007). Leading change-Why transformation efforts fail. *Harvard business review, 85*(1), 96–103.

Kozlowski, S. W. J., Chao, G. T., Grand, J. A., Braun, M. T., and Kuljanin, G. (2013). Advancing multilevel research design: Capturing the dynamics of emergence. *Organizational Research Methods, 16*, 581–615.

Lam, C. F., Spreitzer, G., and Fritz, C. (2014). Too much of a good thing: Curvilinear effect of positive affect on proactive behaviors. *Journal of Organizational Behavior, 35*(4), 530–546.

Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review, 24*(4), 691–710.

Lempiälä, T. (2011, July 1st). Corporate smugglers: Championing ideas under the radar. *Fast Company.*

Liedtka, J. (2009). Leading Innovation at Kelvingrove (A). *University of Virginia Case Series.*

Markham, S. K., Green, S. G., and Basu, R. (1991). Champions and antagonists: Relationships with Rand D project characteristics and management. *Journal of Engineering and Technology Management, 8*(3/4), 217–242.

Miller, P., and Wedell-Wedellsborg, T. (2013). The Case for Stealth Innovation. *Harvard business review, 91*(3), 90–97.

Miron-Spektor, E., Erez, M., and Naveh, E. (2011). The effect of conformist and attentive-to-detail members on team innovation: Reconciling the innovation paradox. *Academy of Management Journal, 54*(4), 740–760.

Mørk, B. E., and Hoholm, T. (2016). From breakthroughs in knowledge to integration in medical practices. In M. Škerlavaj, M. Černe, A. Carlsen, and A. Dysvik (Eds.), *Capitalizing on*

Creativity at Work: Fostering the Implementation of Creative Ideas in Organizations. Cheltenham, UK: Edward Elgar Publishing.

Nurmohamed, S., and Harrison, S. (2016). *Hearing Crickets: An Inductive Study of Overcoming Negative Reactions to Radical Creativity*.

Ohly, S., and Fritz, C. (2010). Work characteristics, challenge appraisal, creativity, and proactive behavior: A multi?level study. *Journal of Organizational Behavior, 31*(4), 543–565.

Osterwalder, A., and Pigneur, Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers*. John Wiley and Sons.

Parker, S. K., Bindl, U. K., and Strauss, K. (2010). Making things happen: A model of proactive motivation. *Journal of management, 36*(4), 827–856.

Perry-Smith, J. E., and Mannucci, P. V. (2017). From Creativity to Innovation: The Social Network Drivers of the Four Phases of the Idea Journey. *Academy of Management Review, 42*(1), 53–79. doi: 10.5465/amr.2014.0462

Phillips, D. J., and Zuckerman, E. W. (2001). Middle-status conformity: Theoretical restatement and empirical demonstration in two markets. *American Journal of Sociology, 107*(2), 379–429.

Rauth, I., and Nabergoj, A. S. (2016). Design thinking workshops: a way to facilitate sensemaking and idea development across organizational levels. In M. Škerlavaj, M. Černe, A. Carlsen, and A. Dysvik (Eds.), *Capitalizing on Creativity at Work: Fostering the Implementation of Creative Ideas in Organizations*. Cheltenham, UK: Edward Elgar.

Rittel, H. W., and Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences, 4*(2), 155–169.

Schön, D. A. (1963). Champions for radical new inventions. *Harvard business review, 41*(2), 77–86.

Simonton, D. K. (2004). *Creativity in science: Chance, logic, genius, and zeitgeist.* Cambridge University Press.

Škerlavaj, M. (2016). Prososial motivasjon: du blir inspirert av gode gjerninger overfor andre … og deg selv. In R. Buch, A. Dysvik, and B. Kuvaas (Eds.), *Produktiv motivasjon i arbeidslivet* (pp. 194–213): Cappelen Damm Akademisk.

Škerlavaj, M., Černe, M., and Dysvik, A. (2014). I get by with a little help from my supervisor: Creative-idea generation, idea implementation, and perceived supervisor support. *The Leadership Quarterly, 25*(5), 987–1000.

Škerlavaj, M., Dysvik, A., Černe, M., and Carlsen, A. (2016). Succeeding with capitalizing on creativity: an integrative framework. In *Capitalizing on Creativity at Work: Fostering the Implementation of Creative Ideas in Organizations* (pp. 335). Cheltenham: Edward Elgar.

Sonenshein, S., and Dholakia, U. (2012). Explaining employee engagement with strategic change implementation: A meaning-making approach. *Organization Science, 23*(1), 1–23.

Stone, G., Russell, R. F., and Patterson, K. (2004). Transformational versus servant leadership: A difference in leader focus. *Leadership and Organization Development Journal, 25*(4), 349–361.

Taras, V., Steel, P., and Kirkman, B. L. (2016). Does country equate with culture? Beyond geography in the search for cultural boundaries. *Management International Review, 56*(4), 455–487.

United Nations, U. (2015). *Transforming our world: the 2030 Agenda for Sustainable Development*. Retrieved from https://sustainabledevelopment.un.org/post2015/transformingourworld

Van de Ven, A. (1999). *The innovation journey*. New York, NY: Oxford University Press.

Van Dierendonck, D. (2011). Servant leadership: A review and synthesis. *Journal of management, 37*(4), 1228–1261.

Van Quaquebeke, N., and Felps, W. (2016). Respectful inquiry: A motivational account of leading through asking open questions and listening. *Academy of Management Review*, amr. 2014.0537.

DOI: 10.18261/9788215031583-2018-13

Chapter 12
The Relative Efficiency of Extrinsic and Intrinsic Motivation

BÅRD KUVAAS

ABSTRACT In this chapter, I review research on the consequences of extrinsic and intrinsic motivation, and conclude that intrinsic motivation is a far more efficient type of motivation than is extrinsic motivation. I also review research on the Job Characteristics Theory (JCT) and Self-Determination Theory (SDT) to pinpoint the most important antecedents to intrinsic motivation, and to show that highly performance-contingent pay can have a detrimental effect on intrinsic motivation, also outside the laboratory.

KEYWORDS: Pay | financial rewards | basic psychological needs | intrinsic motivation | extrinsic motivation

12.1 INTRODUCTION

Motivation can be defined as "energetic forces that initiate work-related behavior and determine its form, direction, intensity and duration" (Pinder, 2008, p. 11). A recent meta-analysis shows that ability and motivation are similarly important to job performance, and that high levels of motivation can compensate for lower levels of ability (Van Iddekinge, Aguinis, Mackey, & DeOrtentiis, 2018). Most, if not all managers understand the importance of having a motivated workforce, but many think about work motivation as a unidimensional construct where the strength of the motivation is the key. In addition, managers have their own personal theories or assumptions about human motivation. McGregor (1960), for instance, argued that managers differed in their assumptions about an average employee at work. Theory X-managers assume that an average employee dislikes work and attempts to avoid work, needs direction, avoids responsibility, lacks ambition, and is motivated by self-interest and maximizing of income. Theory Y-

managers, however, assume that an average employee likes to work, has self-control and direction, seeks responsibility and is motivated to work well and to develop skills and abilities, and desires to participate in tasks that advance worthy organizational goals. Furthermore, Heath (1999) documented through three laboratory studies and one field study that people use lay theories that imply that they have an extrinsic incentives bias. People typically think that others are more motivated by extrinsic incentives such as pay and bonuses than they themselves are. At the same time, people think that others are less motivated by intrinsic "incentives" such as liking and enjoying work and verbal rewards than themselves. Such an extrinsic incentives bias may lead managers to engage in transactional and inefficient leadership behavior, and organizations to implement human resource management (HRM) practices that fail to motivate employees to desirable behaviors. Therefore, in order for managers and organizations to increase employee productivity and well being and organizational productivity, empirical research on employee motivation is crucial.

The chapter is organized as follows. First, I review research on the relative efficiency of extrinsic and intrinsic motivation. In the next main section, I describe the two dominating theoretical frameworks—Job Characteristics Theory (JCT) and Self-Determination Theory (SDT)—and empirical research within these frameworks to demonstrate the most important antecedents to intrinsic motivation. I then review research on the relationship between financial reward and both extrinsic and intrinsic motivation, before ending the chapter with a section on practical implications.

12.2 THE RELATIVE EFFICIENCY OF EXTRINSIC AND INTRINSIC MOTIVATION

So-called instrumentality theories of motivation have also treated motivation as a unidimensional construct, and relied exclusively on extrinsic motivation; that is, the desire to perform an activity with the intention of attaining positive consequences and avoiding negative consequences (E. L. Deci & Ryan, 2000). Reinforcement theory, for instance, posits that behaviors followed by a reinforcer (i.e., something that increases the desired behavioral response) are more likely to recur in the future (e.g. Stajkovic & Luthans, 2003). Expectancy theory (Vroom, 1964), predicts that individuals will engage in behaviors that are likely to lead to valued outcomes, to the extent that they perceive that they are capable to produce such behaviors. Even one of the most applied theoretical lenses in macro HRM research, the ability, motivation and opportunity (AMO) model, relies exclusively

on extrinsic motivation and does not take into account other types of motivation, such as intrinsic and prosocial motivation. Contemporary research knowledge, however, suggests the form of motivation may be more important than its strength. Although there exist numerous constructs to reflect different types of motivation, extrinsic motivation, intrinsic motivation, and prosocial motivation cover the most practically relevant forms of motivation at work and many other contexts, such as sports and education. The main topic of this chapter is intrinsic and extrinsic motivation, although prosocial motivation will be briefly discussed at the end of this section.

When extrinsically motivated, employees do something in order to receive something that is of value to them, such as a promotion, a pay increase or a bonus, or to avoid something negative such as a demotion or being laid off. In our own research, we have used the following items to measure extrinsic motivation (Dysvik & Kuvaas, 2013a):

1. If I am supposed to put in extra effort in my job, I need to get extra pay.
2. It is important for me to have an external incentive to strive for in order to do a good job.
3. External incentives such as bonuses and provisions are essential for how well I perform my job.
4. If I had been offered better pay, I would have done a better job.

Intrinsic motivation, in contrast, is the desire to perform an activity for its own sake, so as to experience the pleasure and satisfaction inherent in the activity (E. L. Deci, Connell, & Ryan, 1989). Thus, intrinsic motivation comes from performing the task, not the consequences of performing it. We have used the following items to measure intrinsic motivation (Kuvaas & Dysvik, 2009):

1. The tasks that I do at work are themselves representing a driving power in my job.
2. The tasks that I do at work are enjoyable.
3. My job is meaningful.
4. My job is very exciting.
5. My job is so interesting that it is a motivation in itself.
6. Sometimes I become so inspired by my job that I almost forget everything else around me.

Although intrinsic and extrinsic motivation can operate simultaneously, the two types of motivation are typically negatively related, although not strongly so (Dysvik & Kuvaas, 2013b; Kuvaas, Buch, Weibel, Dysvik, & Nerstad, 2017; Kuvaas & Dysvik, 2016). This means that either intrinsic or extrinsic motivation is predominant when employees are performing their tasks (Gagné & Deci, 2005; Weibel, Rost, & Osterloh, 2010). We already know from a meta-analytical study including more than 212,000 participants that intrinsic motivation is moderately to strongly related to work performance (Cerasoli, Nicklin, & Ford, 2014). Intrinsic motivation has also been found to relate strongly to human wellness across different domains (see Ryan & Deci, 2017 for a review). We have summarized research from organizations located in Norway including more than 11,000 employees and found that intrinsic motivation is, in addition to being positively related to work performance, positively related to organizational citizenship behaviors[1] and affective organizational commitment,[2] and negatively related to turnover intention and stress (Kuvaas & Dysvik, 2016).

In the remaining part of this section, I will review research that has investigated both intrinsic and extrinsic motivation to answer the question of whether one or the other is a more efficient form of motivation.

Relatively few studies have investigated both extrinsic and extrinsic motivation, but across several samples of sales people including more than 3,000 respondents, we found that intrinsic motivation was more than three times stronger positively related to work performance than extrinsic motivation, and more than twice as strongly positively related to organizational citizenship behaviors than extrinsic motivation (Kuvaas & Dysvik, 2016). In addition, intrinsic motivation was positively related to affective organizational commitment, whereas there was no relationship between extrinsic motivation and affective organizational commitment. Finally, extrinsic motivation was positively and intrinsic motivation negatively related to turnover intention. Thus, extrinsically motivated employees also think more about quitting, whereas intrinsically motivated employees think less about it.

In a recent study, we investigated the relative efficiency of intrinsic and extrinsic motivation in three different samples (Kuvaas, Buch, et al., 2017). In the first

1. Organizational citizenship behaviors are positive behaviors that are discretionary and not directly recognized by the formal reward system in the organization (Organ, 1997).
2. Affective organizational commitment refers to an affective or emotional attachment to the organization such that employees identify with, are involved in, and enjoy membership in the organization (Allen & Meyer, 1990). It is the component of commitment that is most strongly related to desirable work behaviors (Meyer, Stanley, Herscovitch, & Topolnytsky, 2002).

sample, consisting of 552 employees working in gas stations, intrinsic motivation was positively related, and extrinsic motivation negatively related, to supervisor-rated work performance after it was statistically controlled for tenure, gender, and the geographic location of the gas stations. In a cross-lagged study of more than 4,500 employees in the financial industry, intrinsic motivation at time 1 was associated with higher affective commitment and lower continuance commitment[3], turnover intention, burnout, and work-family conflict at time 2. Extrinsic motivation at time 1 was, in contrast, associated with lower affective commitment and higher continuance commitment, turnover intention, burnout, and work–family conflict at time 2. This sample was controlled for several potentially competing variables such as gender, education, tenure, employment conditions, managerial responsibility, and pay level. The third and final sample consisted of 829 employees from a medical technology company and a company in the financial industry. In this sample, extrinsic motivation was unrelated to supervisor-rated work performance, but positively related to turnover intention, whereas intrinsic motivation was positively associated with supervisor-rated work performance and negatively associated with turnover intention. Thus, across three samples the most positive finding for extrinsic motivation is that it was unrelated to work performance in one of the samples.

In a sample of 885 Belgian employees, Vansteenkiste et al. (2007) found that an extrinsic work value orientation, which is similar to extrinsic motivation, was associated with lower job satisfaction, life satisfaction, and life happiness after it was controlled for age, education, pay, and overall work value orientation. In another sample, an extrinsic work value orientation was relative to an intrinsic work value orientation associated with lower dedication, job vitality and job satisfaction, and higher short-lived satisfaction, work–family conflict, emotional exhaustion and turnover intention.

Finally, in a decade-long study of over 10,000 cadets by Wrzesniewski et al. (2014) linking intrinsic and extrinsic motives to attend West Point Academy and career success, the stronger the extrinsic motives to attend West Point, the lower the probability of being selected for consideration for early promotion during the five years of mandatory service, and the lower the probability of remaining military officers beyond the mandatory five year period. The opposite was found for intrinsic motives. In addition, the stronger intrinsic motives to attend, the higher the probability of completing studies and of becoming commissioned officers.

3. Continuance commitment denotes the perceived costs associated with leaving the organization and is a component of commitment which is mainly associated with undesirable work behaviors (Meyer et al., 2002).

In addition to these studies, a number of studies that have investigated four to six subtypes of motivation found that the more autonomous motivation (where intrinsic motivation is the most autonomous), the better work performance and well-being at work (see E. L. Deci, Olafsen, & Ryan, 2017 for a review). The more controlled motivation (where extrinsic motivation is the most controlled), in contrast, the lower work performance and well-being, with a few exceptions that may be explained by problems with the measurement instruments (Kuvaas, Buch, et al., 2017).

Prosocial motivation refers to the desire to benefit others and has been found to be associated with higher levels of persistence, work performance, and organizational citizenship behaviors (Grant, 2008). Prosocial motivation is different from intrinsic motivation on three important accounts. First, intrinsic motivation is more autonomous and based on interest and enjoyment, whereas prosocial motivation is based more on conscious self-regulation and self-control to achieve a goal, such as a felt obligation to reciprocate or pay back favorable treatment from others. Secondly, when employees are intrinsically motivated, the work is seen as an end in and of itself, but when employees are prosocially motivated, they see the work as a means to the end goal of benefiting others. Third, when employees are intrinsically motivated, they focus on the work here and now, but when they are prosocially motivated, they are more oriented toward achieving a goal in the future. These differences make intrinsic motivation more pleasant and joyful than prosocial motivation, and Grant has found that intrinsic motivation reinforces the positive implications of prosocial motivation (Grant, 2008). For instance, the relationship between prosocial motivation, task persistence and work performance are strengthened when intrinsic motivation is high.

That intrinsic motivation can serve as a reinforcer has received relatively broad support. I found, for instance, a positive relationship between performance appraisal satisfaction and work performance for employees with high intrinsic motivation (Kuvaas, 2006a). For employees with low intrinsic motivation, there was a negative association between performance appraisal satisfaction and work performance. In another study, we found a positive relationship between perceived training opportunities and organizational citizenship behaviors only for highly intrinsically motivated employees (Dysvik & Kuvaas, 2008). Furthermore, we only found a positive relationship between perceived investment in employee development and organizational citizenship behaviors for employees with high levels of intrinsic motivation (Kuvaas & Dysvik, 2009). Finally, we found that highly intrinsically motivated employees shared knowledge independently of perceived training intensity, whereas employees with low intrinsic motivation only shared knowledge when pushed by high levels of perceived training intensity (Kuvaas, Buch, & Dysvik, 2012).

In conclusion, intrinsic motivation is a strong predictor of productive work behaviors and attitudes. In addition, intrinsic motivation seems to reinforce the consequences of factors such as prosocial motivation and perceptions of HR practices. Nevertheless, I do not suggest that extrinsic motivation cannot positively influence work performance. Extrinsic motivation can be effective in influencing work performance where there is little potential for intrinsic motivation, such as for simple and uninteresting tasks (Weibel et al., 2010), and when it is relatively easy to measure and monitor work outcomes. Such tasks, however, can often be automatized or performed in countries with cheaper labor.

12.3 ANTECEDENTS TO INTRINSIC MOTIVATION

Below, I will first present the dominant conceptual frameworks that have been applied to understanding the antecedents to intrinsic motivation more deeply. Then, I will review research investigating how financial rewards are related to extrinsic and intrinsic motivation.

12.3.1 THEORETICAL FRAMEWORKS AND ANTECEDENTS

The two most dominating conceptual frameworks applied in research on intrinsic motivation have been the Job Characteristics Theory (JCT) (Oldham & Hackman, 2010) and Self-Determination Theory (SDT) (Gagné & Deci, 2005). JCT posits that five core job characteristics (as perceived by the employee) predict intrinsic motivation (and other outcomes). JCT has received substantial empirical support from several meta-analyses. In the latest, with almost 220,000 respondents, all five job characteristics are moderately to strongly positively related to intrinsic motivation. The five job characteristics are:

- Skill variety
  - The extent to which the tasks performed at work require a variety of different activities, involving the use of a number of different competencies.
- Task identity
  - The degree to which the tasks performed at work require doing an identifiable and whole piece of work.
- Task significance
  - The extent to which the tasks performed at work have an impact on the lives of other people inside or outside the organization.

- Autonomy
  - The degree to which the tasks performed at work provide freedom, independence, and discretion to the employee in scheduling the work and in deciding the procedures and activities to be used at work.
- Feedback from the job
  - The extent to which the tasks performed at work provide the employee with feedback about the effectiveness of his or her performance.

In SDT it is argued that satisfaction of three basic psychological needs for autonomy, competence, and relatedness are essential for individuals to achieve psychological growth, which is manifested by intrinsic motivation (E. L. Deci & Ryan, 2000). A recent meta-analysis of more than 45,000 employees shows that satisfaction of these needs are from moderately to strongly positively related to intrinsic motivation, where the far strongest predictor is satisfaction of the need for autonomy (Van den Broeck, Ferris, Chang, & Rosen, 2016). A closer description of the needs is provided below:

- The need for autonomy
- The need for autonomy refers to the assumption that individuals' need to act with a sense of ownership of their behavior, and to feel psychologically free (Deci & Ryan, 2000). It implies that individuals function at their best when they are the origin of their own behaviors, rather than being regulated by external forces. The need for autonomy should not be confused with a need to behave independently from others. It does, however, imply a need to act with a sense of choice and volition. If a manager asks an employee to work during the weekend and the employee wants to do so, the need for autonomy will be satisfied. If the manager demands that the employee should work during the weekend and the employee wants to do something else, the need for autonomy will be thwarted. The need for autonomy is the most important need, and a necessity that must be fulfilled in order to experience intrinsic motivation (Ryan & Deci, 2006).
- The need for competence
  - The need for competence refers to the need to feel a sense of mastery over the environment and to develop new skills (Van den Broeck et al., 2016). When the need for competence is satisfied, individuals act to maintain and develop their skills, as well as seek challenges that are optimal for their capacity (Elliot, McGregor, & Thrash, 2002).

- The need for relatedness
  - The need for relatedness refers to the need to feeling connected to others, and caring for and being cared for by others, as well as having a sense of belongingness to others such as groups, communities or organizations (Ryan & Deci, 2002). When the need for relatedness is satisfied, individuals feel safe in a psychological sense and feel free to express their personal or work-related challenges (Baumeister & Leary, 1995)

12.3.2 FINANCIAL REWARDS AND EXTRINSIC AND INTRINSIC MOTIVATION

For almost five decades, there has been a fierce debate about whether financial rewards can undermine intrinsic motivation. According to SDT, the effect of financial rewards on motivation depends on two contrasting effects on employees' locus of control: an informing effect that facilitates the feeling of competence and autonomy, which increases intrinsic motivation, and a controlling effect, which undermines intrinsic motivation (E. L. Deci et al., 2017). Pay with low performance contingency, such as base pay and pay that comes as a surprise, is predicted to increase intrinsic motivation because it can inform the recipients about their level of competence and be interpreted as a token of appreciation for good performance. In addition, the base is guaranteed, given prior to future performance. This may signal trust in employees, as they do not have to meet specific future performance standards in order to receive it and, therefore, satisfy the need for competence. Furthermore, a comparably high base pay level sends a signal of how much the organization values employees for what they are (Gardner, Van Dyne, & Pierce, 2004; Kuvaas, 2006b). Pay that is highly contingent on future specific performance standards or results, such as bonuses and commissions, however, are predicted to strengthen employees' perceptions of an external locus of control and, therefore, to reduce autonomy and intrinsic motivation. If an employee does something in order to receive a bonus that he or she would not have done without the opportunity to receive a bonus, the employee is controlled by pay, and the satisfaction of the need for autonomy is thwarted, which will increase extrinsic motivation. Bonuses and commission also put more of the employees' pay at risk and create an economic exchange relationship with the organization (Kuvaas, Shore, Buch, & Dysvik, 2017). Proponents of individual variable pay, on the other hand, view high performance contingency as a necessary condition for "sufficiently strong" incentive effects (Gerhart, 2017), where incentive effect is the degree to which the reward actually changes behaviors.

Most research on the undermining effect of financial rewards has been conducted in the laboratory. A meta-analysis of 128 experimental studies shows that tangible rewards have a negative effect on intrinsic motivation for interesting tasks among children and college students (E. L. Deci, Ryan, & Koestner, 1999). Verbal rewards or positive feedback, in contrast, were found to increase intrinsic motivation. In addition, a vignette study with executive MBA students also shows that financial rewards can undermine intrinsic motivation (Weibel et al., 2010). Such findings cannot, however, be extrapolated to the organizational field because of differences in importance, size, and time frame between laboratory rewards studies and real world compensation systems (Gagné & Forest, 2008; Gerhart & Fang, 2015).

Despite the ongoing debate about rewards and motivation, few field studies have directly addressed the matter. Indirect support for an undermining effect of financial rewards, however, is found in the mentioned meta-analyses by Cerasoli et al. (2014). They found that intrinsic motivation was stronger related to performance when rewards with low performance contingency were present (e.g. base pay or collective rewards)—and that intrinsic motivation was weaker related to performance when rewards with high performance contingency were present (e.g. individual variable incentives such as bonuses and commissions). These findings imply that it is not the presence of rewards that matter, but the performance contingency of the rewards.

Cerasoli et al. (2014) did not investigate extrinsic motivation, nor did the studies in the meta-analysis include actual pay data. In the first study that simultaneously investigated actual pay data from different pay components and both intrinsic and extrinsic motivation, we found that the amount of base pay received over two years was positively related to intrinsic motivation among sales people in an insurance company (Kuvaas, Buch, Gagné, Dysvik, & Forest, 2016). In a study among knowledge workers, I also found a positive relationship between base pay level and intrinsic motivation (Kuvaas, 2006b), supporting an informing effect of pay with low performance contingency. The amount of money received as annual bonuses over the two years, in contrast, was negatively related to intrinsic motivation and positively related to extrinsic motivation, supporting a controlling effect of pay with high performance contingency. Furthermore, intrinsic motivation was strongly related to an increase in work effort and strongly related to a decrease in turnover intention. Extrinsic motivation was weakly related to an increase in work effort (.11 compared to .61 for intrinsic motivation), but positively related to an increase in turnover intention. The weak positive relationship between extrinsic motivation and increase in work effort is, thus, more than out-

weighed by the negative relationship between the amount of money received as annual bonuses and intrinsic motivation, and the positive association between extrinsic motivation and increase in turnover intention. These findings, and others revealed in this chapter, strongly refute Gerhart and Fang (2014, p. 47), who argued that "if there is an undermining effect on intrinsic motivation, it is usually dominated by the positive effect of PFIP (pay-for-individual-performance) on extrinsic motivation."

To conduct a more explicit test of the informing and controlling effects of financial rewards, we recently developed a measure of incentive effects that assesses the degree to which a pay plan affects employee behaviors (Kuvaas, Buch, & Dysvik, 2018). Example items are "The pay plan makes me do things I would not have done if we did not have the pay plan we have," "The pay plan affects my daily priorities," and "If we did not have the pay plan we currently have, I would have performed my tasks at work in a different way." The most prevalent pay plans in the United States (Gerhart, 2017), and probably also elsewhere, are those based on subjective performance evaluations by managers. Because such evaluations are typically based on a large number of often vague criteria, it is difficult for employees to know how to get good evaluations. Therefore, the incentive effects are uncertain. Thus, the degree to which such plans have incentive effects, and how they influence motivation, are important empirical questions. In our study, we investigated 304 employees working at the national headquarters of a retail organization across eight European countries. We found relative weak incentive effects (an average of 2.30 on a scale from one to five) and a weak relationship between the amount of money received as bonuses and incentive effects. The incentive effect, however, was strongly positively related to extrinsic motivation and moderately negatively related to intrinsic motivation, supporting the controlling effect posited by SDT. Accordingly, the pay plan did not work very well, i.e. did not change employee behaviors, but to the extent that it did, the results were higher extrinsic and lower intrinsic motivation, which was not beneficial to the organization. In support of an informational effect, we also found a direct positive association between the amount of money received as bonuses and intrinsic motivation. Thus, when the pay plan did not change the behaviors, it was positively related to intrinsic motivation. In sum, however, the negative association between the incentive effect and intrinsic motivation outweighed the positive association between the reward and intrinsic motivation.

In conclusion, there is sufficient evidence to conclude that highly performance-contingent pay can have a detrimental effect on intrinsic motivation. Less performance-contingent pay, such as a comparably high base pay, however, can increase

intrinsic motivation, also outside the laboratory. The consequences of variable pay with low performance contingency and small or no incentive effects are more difficult to predict.

12.4 PRACTICAL IMPLICATIONS

Overlearned carrot-and-stick and command-and-control approaches have dominated businesses for much more than a century, and Stone, Deci, and Ryan (2009) have proposed six practical steps or actions that facilitate intrinsic motivation as an alternative to "traditional" management.

- Asking open questions, including inviting participation, in solving important problems
  - Facilitating intrinsic motivation requires supportive dialogue, and supportive dialogue starts with open questions that invite exploration of an important problem. Open questions such as "What do you make of this?" or "Can you enlighten me on the current project?" raise issues for consideration without implying a preferred solution. Closed questions, in contrast, such as "Have you tried to fix the problem by …?" imply a need for passive compliance.
- Active listening, including acknowledging the employees' perspective
  - Open questions should be followed by active and reflective listening that explicitly acknowledges the employee's perspective or perception of a situation. This requires careful attention to the emotional aspects of an issue and an empathetic leadership style.
- Offering choices within structure, including the clarification of responsibilities
  - As the satisfaction of the need for autonomy and the perception of job autonomy are the most important predictors of intrinsic motivation, offering a menu of alternative actions to deal with a situation is crucial. This is especially important when faced with problems (e.g. cost cutting) in addition to providing a meaningful rationale (e.g. for a boring, but important task).
- Providing sincere, positive feedback that acknowledges initiative, and factual, non-judgmental feedback about problems
  - Verbal rewards can facilitate intrinsic motivation by supporting competence and autonomy. Effective feedback is specific, sincere, and acknowledges unique contributions. Verbal rewards should also be provided for proactive behaviors and initiatives. Corrective feedback should be constructive and task focused, not person focused, and be provided in a timely fashion close to the behavior (Kuvaas, Buch, & Dysvik, 2016).

- Minimizing coercive controls such as rewards and comparisons with others
  - As we have seen, pay with high performance contingency and incentive effects should be avoided if organizations want to facilitate intrinsic motivation. A general advice is to pay employees adequately and fairly and minimize the salience of compensation; that is, to get the money "off the table." In addition to performance contingent financial rewards, SMART (specific, measurable, attainable, realistic, and timely/tangible) goals (Kuvaas & Buch, 2017; Kuvaas, Buch, & Dysvik, 2014), deadlines (Amabile, DeJong, & Lepper, 1976), evaluations (Smith, 1975), and surveillance (Lepper & Greene, 1975) have also been found to relate negatively to intrinsic motivation or need satisfaction. Comparing employees to each other and fostering internal competition is also detrimental to intrinsic motivation. Accordingly, ranking and rating performance management systems that were widely used before Microsoft abandoned theirs in 2013 should clearly be avoided.
- Develop employees and share knowledge to enhance competence and autonomy
  - The motivational consequences of training and development opportunities and promotions depend on why employees desire them, and how and why they are offered. Such opportunities should therefore be provided as a means to provide more autonomy, learning new skills, and collaborating with others to satisfy the needs for autonomy, competence, and relatedness. If training and development opportunities and promotions are offered in an "if-you-do-this-then-you-get-that" fashion, they may decrease intrinsic and increase extrinsic motivation.

When applying these steps or actions it is important to acknowledge that individual employees may have different levels of the three needs—based on, for instance, their experience and personality. Employees who have a very high need for competence should get more frequent verbal praise than others, whereas those with a high need for autonomy should be provided more discretion, choice, and autonomy than others.

In addition to the steps or actions described above, it is important to expand jobs to provide optimal challenges and facilitate participation and involvement in accordance with JCT. Broad job tasks with skill flexibility, job variety, and autonomy, and responsibility that focuses on the purpose of the job, increases intrinsic motivation. Narrowly defined jobs with detailed and specific descriptions and instructions about tasks and behaviors, in contrast, are detrimental to intrinsic motivation. Moreover, since employees may have different needs, interests, and

strengths, the same titles or jobs do not need to be designed in exactly the same way. Effective job design therefore also includes taking individual differences into account.

With respect to recruitment, the meta-analysis by Van den Broeck et al. (2016) shows that demographics such as age, sex, and organizational tenure are very weakly related to satisfaction of the needs for autonomy, competence, and relatedness, and thereby intrinsic motivation. Personality factors, however, such as optimism, self-esteem and efficacy, agreeableness, and conscientiousness are moderately to strongly positively related to need satisfaction, whereas neuroticism is relatively strongly negatively related to need satisfaction. Accordingly, organizations can to some extent influence need satisfaction and intrinsic motivation through their recruitment processes by recruiting people based on these personality factors.

Job stressors like role conflict, role ambiguity, organizational politics, and job insecurity are negatively associated with need satisfaction (Van den Broeck et al., 2016) implying that downsizing, unclear and unfair practices and procedures should be avoided, and training encouraged. The strongest predictors of need satisfaction are contextual factors such as leadership and perceived organizational support and fairness. Among several leadership factors, high-quality leader–member exchange relationships are very strongly positive related to need satisfaction. Accordingly, leadership training that focuses on soft skills can facilitate need satisfaction and thereby intrinsic motivation (see Lacerenza, Reyes, Marlow, Joseph, & Salas, 2017 for a prescription of effective leadership training).

REFERENCES

Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance, and normative commitments to organization. *Journal of Occupational Psychology, 63*, 1–8.

Amabile, T. M., DeJong, W., & Lepper, M. R. (1976). Effects of externally imposed deadlines on subsequent intrinsic motivation. *Journal of Personality and Social Psychology, 34*, 92–98.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529. doi: 10.1037/0033-2909.117.3.497

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin, 140*(4), 980–1008. doi: 10.1037/a0035661

Deci, E. L., Connell, J. P., & Ryan, R. M. (1989). Self-determination in a work organization. *Journal of Applied Psychology, 74*(4), 580–590.

Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-determination theory in work organizations: The state of a science. *Annual Review of Organizational Psychology and Organizational Behavior, 4*(1), 19–43. doi: 10.1146/annurev-orgpsych-032516-113108

Deci, E. L., & Ryan, A. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227–268.

Deci, E. L., Ryan, R. M., & Koestner, R. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668.

Dysvik, A., & Kuvaas, B. (2008). The relationship between perceived training opportunities, work motivation and employee outcomes. *International Journal of Training and Development, 12*(3), 138–157.

Dysvik, A., & Kuvaas, B. (2013a). Intrinsic and extrinsic motivation as predictors of work effort: The moderating role of achievement goals. *British Journal of Social Psychology, 52*(3), 412–430. doi: 10.1111/j.2044-8309.2011.02090.x

Dysvik, A., & Kuvaas, B. (2013b). Intrinsic and extrinsic motivation as predictors of work effort: The moderating role of achievement goals. *British Journal of Social Psychology, 52*, 412–430.

Elliot, A. J., McGregor, H. A., & Thrash, T. M. (2002). The need for competence. In E. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 361-387). Rochester, NY: The University of Rochester Press.

Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior, 26*, 331-362. doi: 10.1002/job.322

Gagné, M., & Forest, J. (2008). The study of compensation through the lens of self-determination theory: Reconciling 35 years of debate. *Canadian Psychology, 49*(3), 225–232.

Gardner, D. G., Van Dyne, L., & Pierce, J. L. (2004). The effects of pay level on organization-based self-esteem and performance: A field study. *Journal of Occupational and Organizational Psychology, 77*, 307–322.

Gerhart, B. (2017). Incentives and Pay for Performance in the Workplace. In A. J. Elliot (Ed.), *Advances in Motivation Science* (Vol. 4, pp. 91–140): Elsevier.

Gerhart, B., & Fang, M. (2014). Pay for (individual) performance: Issues, claims, evidence and the role of sorting effects. *Human Resource Management Review, 24*, 41–52.

Gerhart, B., & Fang, M. (2015). Pay, intrinsic motivation, extrinsic motivation, performance, and creativity in the workplace: Revisiting long-held beliefs. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 489–521. doi: 10.1146/annurev-orgpsych-032414-111418

Grant, A. M. (2008). Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity. *Journal of Applied Psychology, 93*(1), 48–58.

Heath, C. (1999). On the social psychology of agency relationships: Lay theories of motivation overemphasize extrinsic incentives. *Organizational Behavior and Human Decision Processes, 78*(1), 25–62.

Kuvaas, B. (2006a). Performance appraisal satisfaction and employee outcomes: Mediating and moderating roles of motivation. *International Journal of Human Resource Management, 17*(3), 504–522.

Kuvaas, B. (2006b). Work performance, affective commitment, and work motivation: The roles of pay administration and pay level. *Journal of Organizational Behavior, 27*(3), 365–385.

Kuvaas, B., & Buch, R. (2017). Leader-member exchange relationships and follower outcomes: The mediating role of perceiving goals as invariable. *Human Resource Management*, 1–14. doi: 10.1002/hrm.21826

Kuvaas, B., Buch, R., & Dysvik, A. (2012). Perceived training intensity and knowledge sharing: Sharing for intrinsic and prosocial reasons. *Human Resource Management, 51*(2), 167–187. doi: 10.1002/hrm.21464

Kuvaas, B., Buch, R., & Dysvik, A. (2014). Performance management: Perceiving goals as invariable and implications for perceived job autonomy and work performance. *Human Resource Management, 55*(3), 401–412. doi: 10.1002/hrm.21680

Kuvaas, B., Buch, R., & Dysvik, A. (2016). Constructive supervisor feedback is not sufficient: Immediacy and frequency is essential. *Human Resource Management*. doi: 10.1002/hrm.21785

Kuvaas, B., Buch, R., & Dysvik, A. (2018). Individual variable pay for performance, incentive effects, and employee motivation. *Accepted for presentation at the annual meeting of the Academy of Management, Chicago, USA*.

Kuvaas, B., Buch, R., Gagné, M., Dysvik, A., & Forest, J. (2016). Do you get what you pay for? Sales incentives and implications for motivation and changes in turnover intention and work effort *Motivation and Emotion, 40*(5), 667–680. doi: 10.1007/s11031-016-9574-6

Kuvaas, B., Buch, R., Weibel, A., Dysvik, A., & Nerstad, C. G. L. (2017). Do intrinsic and extrinsic motivation relate differently to employee outcomes? *Journal of Economic Psychology, 61*(244–258). doi: 10.1016/j.joep.2017.05.004

Kuvaas, B., & Dysvik, A. (2009). Perceived investment in employee development, intrinsic motivation, and work performance. *Human Resource Management Journal, 19*(3), 217–236.

Kuvaas, B., & Dysvik, A. (2016). *Lønnsomhet gjennom menneskelige ressurser: Evidensbasert HRM* (3. utgave ed.). Oslo: Fagbokforlaget.

Kuvaas, B., Shore, L. M., Buch, R., & Dysvik, A. (2017). Social and economic exchange relationships and performance contingency: differential effects of variable pay and base pay. *The International Journal of Human Resource Management,* 1–24. doi: 10.1080/09585192.2017.1350734

Lacerenza, C. N., Reyes, D. L., Marlow, S. L., Joseph, D. L., & Salas, E. (2017). Leadership training design, delivery, and implementation: A meta-analysis. *Journal of Applied Psychology, 102*(12), 1686–1718. doi: 10.1037/apl0000241

Lepper, M. R., & Greene, D. (1975). Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Journal of Personality and Social Psychology, 31*, 479–486.

McGregor, D. (1960). *The human side of enterprise*. New York: McGraw-Hill.

Meyer, J. P., Stanley, D. J., Herscovitch, L., & Topolnytsky, L. (2002). Affective, continuance, and normative commitment to the organization: A meta-analysis of antecedents, correlates, and consequences. *Journal of Vocational Behavior, 61*, 20–52.

Oldham, G. R., & Hackman, J. R. (2010). Not what it was and not what it will be: The future of job design research. *Journal of Organizational Behavior, 31*(2-3), 463–479. doi: 10.1002/job.678

Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance, 10*, 85–97.

Pinder, W. C. C. (2008). *Work motivation in organizational behavior* (2nd ed.). New York: Psychology Press.

Ryan, R. M., & Deci, E. (2002). An overview of self-determination theory: An organismic-dialectical perspective. In E. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 3–36). Rochester, NY: The University of Rochester Press.

Ryan, R. M., & Deci, E. L. (2006). Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of Personality, 74*(6), 1557–1585.

Ryan, R. M., & Deci, E. L. (2017). Self-determination theory: *Basic psychological needs in motivation, development, and wellness*. New York: Guilford.

Smith, W. E. (1975). *The effect of anticipated vs. Unanticipated social reward on subsequent intrinsic motivation*. Unpublished dissertation, Cornell University.

Stajkovic, A. D., & Luthans, F. (2003). Behavioral management and task performance in organizations: Conceptual backgrounds, meta-analysis, and test of alternative models. *Personnel Psychology, 56*, 155–194.

Stone, D. N., Deci, E. L., & Ryan, R. M. (2009). Beyound talk: Creating autonomous motivation through self-determination theory. *Journal of General Management, 34*(3), 75–91.

Van den Broeck, A., Ferris, D. L., Chang, C.-H., & Rosen, C. C. (2016). A review of self-determination theory's basic psychological needs at work. *Journal of Management, 42*(5), 1195–1229.

Van Iddekinge, C. H., Aguinis, H., Mackey, J. D., & DeOrtentiis, P. S. (2018). A meta-analysis of the interactive, additive, and relative effects of cognitive ability and motivation on performance. *Journal of Management, 44*(1), 249–279. doi: 10.1177/0149206317702220

Vansteenkiste, M., Neyrinck, B., Niemiec, C. P., Soenens, B., De Witte, H., & Van den Broeck, A. (2007). On the relations among work value orientations, psychological need satisfaction and job outcomes: A self-determination theory approach. *Journal of Occupational and Organizational Psychology, 80*(2), 251–277. doi: 10.1348/096317906X111024

Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.

Weibel, A., Rost, K., & Osterloh, M. (2010). Pay for performance in the public sector: Benefits and (hidden) costs. *Journal of Public Administration Research and Theory, 20*(2), 387–412.

Wrzesniewski, A., Schwartz, B., Cong, X., Kane, M., Omar, A., & Kolditz, T. (2014). Multiple types of motives don't multiply the motivation of west point cadets. *Proceedings of the National Academy of Sciences of the United States of America, 111*(30), 10990–10995. doi: 10.1073/pnas.1405298111

DOI: 10.18261/9788215031583-2018-14

Chapter 13 Knowledge Spillovers and the Timing of Environmental R&D Subsidies

GEIR H.M. BJERTNÆS, TOM-REIEL HEGGEDAL AND KARL JACOBSEN

ABSTRACT Several recent studies have concluded that subsidies for environmentally friendly R&D should be high initially and decline over time. This study shows that scale aspects connected to knowledge spillovers from environmental R&D support the opposite conclusion. Increasing returns to scale in the production of abatement knowledge, as well as an increasing price of carbon emissions, are aspects that favor increasing subsidy rates to firms conducting environmentally friendly R&D.

JEL classification: O32; O38.

KEYWORDS: Environment | Innovation policy | Knowledge spillovers

13.1 INTRODUCTION

The goal of the Paris Agreement is 'holding the increase in global average temperature to well below 2°C above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5°C' (UNFCCC, 2015). A major challenge for this goal is to reduce the share of fossil fuels in the energy mix. To keep global warming below 2°C, a third of oil reserves, a half of gas reserves, and more than 80 percent of coal reserves must stay in the ground (McGlade and Ekins, 2015).[1] On the

1. In a recent paper, Milllar et al. (2017) are more positive on how much fossil energy can be used to still stay within the target. They show that by limiting cumulative post-2015 CO_2-emissions to about 750 bn tonnes of CO_2, the post-2015 global warming can be limited to less than 0.6°C (temperature increase from 1870 till today is about 0.9°C). Annual emissions at the moment are almost 40 bn tonnes of CO_2, which implies that we will exceed the 1.5°C carbon budget in 19 years if we stay on the current path.

other hand, the IEA (2017) predicts a 30% growth in total energy demand between today and 2040. Hence, the production of clean energy must increase dramatically.

Research and development (R&D) drives down costs and improves technologies, and hence facilitates the diffusion of new, clean technologies. Such technological improvements are considered to be a key element in curbing global warming (see e.g., Carraro et al., 2003; Jaffe et al., 2005). However, innovation markets are inefficient and it is well known that the social returns to R&D may be greater than the private returns to R&D (see e.g. Griliches, 1995; Jones and Williams, 2000; Klette et al., 2000; Bloom et al., 2013). A major reason for this is the public-good nature of knowledge. Knowledge generated from a firm's R&D activity might spill over to other firms and expand future R&D opportunities, i.e., when developing new ideas we are "standing on the shoulders of giants" (Isaac Newton). The value of these knowledge spillovers are only partly captured by the individual firms, and R&D activity should be subsidized.

There is a recent literature on climate policy and directed technological change that analyzes the timing of subsidies to R&D in clean technologies. Acemoglu et al. (2012) show that clean innovation should be heavily subsidized in early periods and decline over time to induce a switch from dirty to clean technologies. The reason is that the subsidy to clean technologies is used to deal with future environmental externalities when patents have a short lifetime.[2] Greaker, Heggedal and Rosendahl (2017) back up the main conclusion on the subsidy path to clean innovation in a model where patents are long-lived and the carbon emission tax perfectly internalizes emission damages. They show that the subsidies to clean innovation should be larger than those to dirty innovation due to knowledge spillovers. Intuitively, climate change necessitates a shift to clean innovation, and when more future researchers innovate on clean technologies than dirty technologies, there will be more researchers standing on the shoulders of a clean innovation done today than of a dirty innovation done today.[3]

We contribute to this literature by analyzing how dynamic aspects of knowledge spillovers, as well as an increasing price of carbon emissions, influence the optimal timing of R&D subsidies to clean technologies, and we find support for the opposite conclusion: Increasing returns to scale in the production of environmental knowledge favor increasing subsidy rates to clean innovation.

We develop a model where the production of new ideas is given by the input of researchers and the stock of knowledge, where the stock of knowledge is a public

2. A similar argument is made by Acemoglu et al. (2016).
3. A similar argument is made by Hart (2018). See Fischer and Heutel (2013) for a survey over the earlier literature on directed technical change in an environmental context.

good.[4] Patent lifetime is infinite and the social value of a patent is assumed to be identical with the patent price, i.e. the R&D firms appropriate the full value in the market of the technologies they develop. The only externality in the model stems from knowledge spillovers, where knowledge spillovers is defined as the productivity effect following from a change in the knowledge stock.[5] A welfare-maximizing government determines R&D subsidy rates each period to harvest potential welfare gains connected to the positive external effects of these spillovers.

The elasticity of scale of the R&D production function turns out to be crucial in our analysis. The reason is that this elasticity captures a combination of how much R&D productivity increases when the knowledge stock grows, and of how much the market responds to this productivity increase. In fact, in an unregulated economy (i.e., zero subsidies) the elasticity of scale determines whether the production of patents increases over time. Model simulations show that optimal R&D subsidy rates increase (decrease) over time when the elasticity of scale is larger (smaller) than one–holding the price of patents constant over time. When choosing subsidies, the government trades off the costs of using more resources on R&D today against the value of knowledge spillovers from expanding patent production. Expanding the production of patents today increases knowledge spillovers through two channels. First, the expansion directly generates a larger stock of knowledge that increases the productivity of future researchers. Second, R&D firms respond by hiring more researchers in the future and, thus, further increase the generation of knowledge. The value of these knowledge spillovers is the market value of the additional patents generated. The market value is given by the patent price, while the extent of increase in patent production depends on a combination of how much R&D productivity increases and of how many researchers the R&D firms hire–which, again, is given by the elasticity of scale of the R&D production function.

Next, we analyze how changes in the price of carbon emissions affect our results. A common result in the literature on climate change is that the optimal tax on carbon emissions should increase over time as the global warming problem escalates, and thus, the social cost of emissions increases.[6] An increasing price of

4. The R&D production sector is similar to R&D production in Romer (1990).
5. There is empirical evidence that environmental R&D to some extent crowds out conventional R&D, and that the social returns to environmental R&D may exceed the social returns to R&D in general (see Popp and Newell, 2012; Dechezleprêtre et al., 2017). The present study abstracts from such crowding out effects, and focuses only on mechanisms related to environmental R&D in isolation.
6. See e.g., Hoel and Kverndokk (1996), Smulders and Bretschger (2000), Nordhaus (2008).

carbon emissions is then incorporated into our model framework by letting the price of environmental patents increase over time. Model simulations show that optimal R&D subsidy rates are increasing over time when the elasticity of scale of the R&D production function equals (or is slightly below) one–when the price of patents increases over time. The reason is that a higher price of patents in a future period implies more future R&D production. When future R&D production is higher, the value of increasing the productivity of R&D goes up, and the government responds by subsidizing R&D more today to build a larger knowledge stock for tomorrow. That is, increasing patent prices exacerbate the knowledge spillovers problem.

Last, model simulations also show that a modest rate of depreciation on the stock of knowledge (due to, for instance, creative destruction), and/or imposing a government budget constraint, do not alter the main results. A drastic rate of depreciation may, however, lead to substantial changes to the above results.

Empirical evidence suggests that it is difficult to pinpoint the elasticity of scale of the environmental R&D production function. Jones (1995, 1999) argues that the return to R&D in general decreases over time as the increase in new knowledge due to a marginal increase in the stock of knowledge is less than one. The scale elasticity, however, also includes the marginal productivity of researchers, and hence, can exceed one even when the marginal returns to the stock of knowledge is less than one. The majority of studies attempting to estimate the environmental R&D production function find that the elasticity of scale exceeds one, see Porter and Stern (2000), Gong et al. (2004), Abdih and Joutz (2005), Pessoa (2005), and Samaniego (2007) for estimates of the elasticity. The future price of carbon emissions is also highly uncertain. However, a large majority of studies predict that the price of carbon emissions will increase as problems connected to global warming escalate. In sum, the empirical evidence, in combination with our simulation results, favors subsidy rates that are increasing over time to firms conducting environmentally friendly R&D.

There are only a few other studies that specifically analyze the timing aspect of spillovers and subsidies to R&D. Grossmann et al. (2010) show that dynamically optimal R&D subsidy rates in general depend on gaps in the stock of knowledge and the capital stock relative to their steady state levels in a model with knowledge spillovers, where the elasticity of scale of R&D production is larger than one, and duplication externalities are included. Heggedal (2015) argues that emerging R&D should be subsidized more heavily compared to mature R&D because the external effect is larger in emerging technologies that grow faster than mature technologies, if the elasticity of scale in R&D production is less than one. Perez-

Sebastian (2007) shows that R&D subsidy rates in general should rise over the transition path towards the balanced growth path within a model with imitation of foreign ideas, as negative externalities connected to imitations gradually decreases.

In an environmental context with increasing carbon taxes, Gerlagh et al. (2009; 2014) show that the optimal R&D subsidy rate to abatement technologies falls over time when patent lifetime is finite. The reason is that there is insufficient support through markets to develop abatement technologies when the price of carbon is low. Heggedal and Jacobsen (2011) find that environmental R&D reforms where subsidy rates are decreasing over time generate the most efficient outcome within a CGE-model of the Norwegian economy.

The paper is organized as follows. The model is laid out in section 13.2, while section 13.3 analyzes the optimal timing of subsidy rates to environmental R&D using simulations. Section 13.4 concludes.

13.2 MODEL

We develop a partial model of the environmental R&D industry to analyze how the externalities from knowledge spillovers affect the optimal distribution of policy incentives across time.[7] We assume infinite patent lifetime. Alternative modelling assumptions, like limited patent lifetime or creative destruction, would give a similar relationship between R&D activity today and R&D productivity in the future as in our model. However, such assumptions would create a wedge between the private value and the social value of patents as R&D firms would no longer appropriate the full value in the market of the technologies they develop. We abstract away from such problems in order to focus on knowledge spillovers as a source of underinvestment in R&D. The price of patents may vary over time as the tax on carbon emissions may vary. We assume that the price of patents is positively correlated with the price of carbon emissions, as the value of environmentally friendly technology is likely to be higher when costs of emissions are high.

The production of new varieties is given by the input of researchers and the stock of knowledge, where the stock of knowledge is treated as a public good. An unlimited supply of labor is available for the R&D-sector at the alternative value, which constitutes a fixed wage rate. The productivity of new research projects differs because some new ideas are better than others. There is free entry of profit-

7. We do not model the mechanisms for spillovers between firms. For a discussion on worker mobility as the source of spillovers, see Heggedal, Moen and Preugschat (2017).

maximizing R&D firms, and the firms ignore that their efforts contribute to increase the knowledge stock.

A model designed to fit a typical industrialized country is solved numerically for empirically relevant parameter values; see Porter and Stern (2000), Gong et al. (2004), Abdih and Joutz (2005), Pessoa (2005), and Samaniego (2007). A welfare-maximizing government determines R&D subsidy rates each period to harvest potential welfare gains connected to positive external effects of knowledge spillovers.

13.2.1 TECHNOLOGY

The aggregate production of new patents (knowledge) in environmentally friendly technologies is given by the type of production function that is used in endogenous growth models with horizontal innovation, e.g. Romer (1990), and with vertical innovation, e.g. Aghion and Howitt (1992):

$$X_t = A_{t-1}^{\phi} L_t^{\lambda}, \tag{1}$$

where X_t is the production of patents, L_t is the labor input, $\lambda \in [0,1)$ is the output elasticity with respect to labor, A_{t-1} is the accumulated patents from previous periods, i.e. the stock of knowledge, and ϕ is the output elasticity with respect to patents, i.e. the spillover parameter. The decreasing returns with respect to labor on an aggregate level are motivated by heterogeneous productivity between research projects in the R&D industry. The spillover parameter reflects the effect of the existing knowledge stock on the production of new patents: $\phi > 1$ implies increasing returns to knowledge, while $\phi < 1$ implies decreasing returns to knowledge. The elasticity of scale in the R&D production function equals $\phi + \lambda$.

The knowledge stock evolves according to

$$A_t = A_{t-1} + X_t = \sum_{i=0}^{t} X_i, \tag{2}$$

where X_0 is the initial stock of knowledge.

13.2.2 ENVIRONMENTAL R&D INDUSTRY

Firms in the environmental R&D industry sell patents at a given price. We assume that the price of patents, P_t, are identical for all firms. This price may increase over time as the price of carbon emissions increases over time.

When a patent is produced, the knowledge embedded in the patent is freely available to other firms in future periods, i.e. the knowledge stock is a public good. The firms do not take into account that their patent production influences the productivity of future R&D. This is the source of the knowledge spillover problem.

Further, in each period there is a continuum of research projects with different productivity. The productivity of new research projects differs, and high productivity projects generate more patents per researcher compared to low productivity projects. Individual firms are endowed with private information about one of the new research projects. In each period the firms decide whether to enter the industry and sell the patent to the given price P_t.

There is free entry into the industry and the least productive firm to enter earns zero profit, i.e. on an industry level there are decreasing returns to labor. The firms take the wage rate w, the unit subsidy rate on labor z_t, and P_t as given, and firms enter the industry until

$$P_t \frac{\partial X_t}{\partial L_t} - (w - z_t) = 0, \tag{3}$$

where $\frac{\partial X_t}{\partial L_t} = \lambda A_{t-1}^{\phi} L_t^{\lambda-1}$ follows from (1). The free entry condition given by (3) can be solved for L_t to get the labor demand in the R&D industry:

$$L_t = \left(\frac{P_t A_{t-1}^{\phi} \lambda}{w - z_t} \right)^{\frac{1}{1-\lambda}} \text{ for all } t. \tag{4}$$

13.2.3 GOVERNMENT

A welfare function is absent within this partial model framework. We, however, assume that production efficiency is part of a welfare maximizing solution, where the alternative production value of labor equals the wage rate, w. The production value of patents in a period equals the price of patents, P_t, multiplied by the number of patents, X_t. Hence, the objective of the government consists of maximizing the present value of all future patents minus the alternative cost of labor allocated to the research sector. The policy tools of the government are restricted to R&D subsidy rates z_t in each period. Subsidy rates are chosen to adjust the market solution of labor allocated to the research sector according to equation (4). The subsequent impact on the production of patents is given by equation (1), which influences the knowledge spillover according to equation (2). The government is fully

aware of this interaction between subsidy rates and impacts on knowledge spillovers.

The government maximization problem is:

$$\max_{\{z_t\}_{t=1}^{\infty}} \sum_{t=1}^{\infty} \frac{1}{(1+r)^{t-1}} \left(P_t X_t - w L_t \right) \tag{5}$$

s.t. $L_t(z_t) > 0$, (1), (2) and (4),

where r is the discount rate, i.e. the interest rate. Note that this objective function equals the sum of the surpluses of the producers and the surplus of the government (which equals the negative value of R&D subsidies). R&D subsidies constitute an income transfer between the public and the private sector. Hence, the subsidy term is cancelled from the objective function of the government. The objective function does not include an expression for the consumer surplus, since the social value of a patent is assumed to be identical with the patent price.

The solution to this maximization problem is found by constructing a numerical model that is solved by computer simulations. The numerical model contains all crucial elements described above. A detailed calibration to a specific economy is omitted because the main mechanisms in the model that determine the results are not affected.

A simple example is constructed to illustrate that the scale elasticity is crucial for the development of the knowledge spillovers over time. Consider the unregulated equilibrium where there are no subsidies, i.e. $z_t = 0$. Rearranging the first order condition from equation (4) when $z_t = 0$ and using (1) gives the production of patents in the unregulated equilibrium as a function of the knowledge stock, A_{t-1}, and the parameters P_t, λ, w and ϕ:

$$X_t = \left(\frac{P_t \lambda}{w} \right)^{\frac{\lambda}{1-\lambda}} A_{t-1}^{\frac{\phi}{1-\lambda}}. \tag{6}$$

We see that the production function (6) is homogenous of degree $\frac{\phi}{1-\lambda}$. If $\phi + \lambda = 1$, then the production function is homogenous of degree 1, implying constant returns to knowledge on production, i.e. constant spillovers over time. If $\phi + \lambda < 1$ ($\phi + \lambda > 1$), then there is decreasing (increasing) returns to knowledge on production, and spillovers decrease (increase) over time. That is, the scale elasticity determines whether the production of patents increases over time, and knowledge spillovers are larger when the future production of patents is larger.

The analytical solution to the government maximization problem is complex because optimal subsidy rates depend on the knowledge stock in all future periods, and vice versa. Thus, we solve the government maximization problem numerically by computer simulations.

13.3 NUMERICAL ANALYSIS

The government maximization problem is solved numerically by computer simulations. Optimal R&D subsidy rates are found by a solver that searches for optimum within the set of feasible combinations of R&D subsidy rates over time. A number of scenarios are simulated to uncover how different parameter combinations ($\phi + \lambda$) influence optimal R&D subsidy rates over time. We present results for the government problem with specific sets of parameter values that turned out to be crucial for the subsidies. The choice of parameter values may lead to convexity in the objective function, which may exclude a numerical solution. A numerical solution for the government's optimization problem is obtained by calibrating the initial production of patents by adjusting the wage rate. The exogenous wage rate is adjusted so that the production of new knowledge equals 2.5 per cent of the stock of knowledge. A list of chosen parameter values is presented in appendix A.

The first set of simulation results is presented in Figure 13.1 for different values of the elasticity of scale in the R&D production function. The price of new patents is assumed to be constant in all future periods in these scenarios.

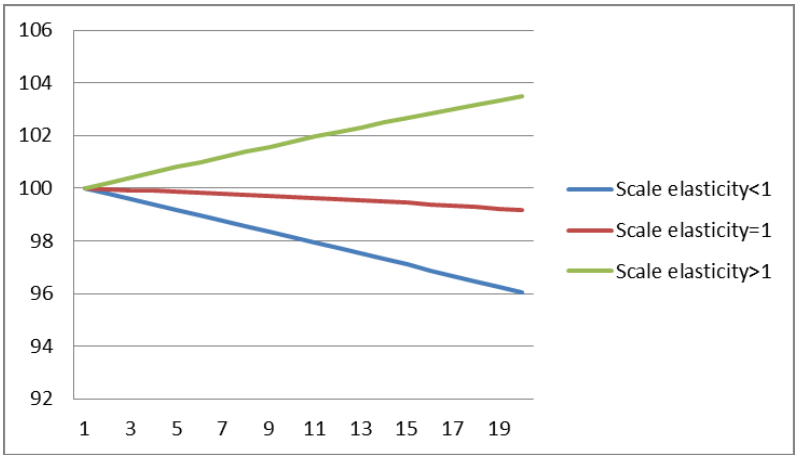

FIGURE 13.1: Optimal subsidy rates with constant patent price

The initial subsidy rate within each scenario is set equal to 100 to facilitate a comparison of scenarios. Figure 13.1 illustrates that the optimal subsidy rates increases (decreases) over time when the elasticity of scale is larger (smaller) than one. Optimal subsidy rates are virtually unchanged when the elasticity of scale equals one. The intuition is that R&D subsidies, which contribute to expand the production of patents, generate two effects. First, the expansion in the production of patents generates a larger future stock of knowledge. Second, the larger future stock of knowledge contributes toward increasing the productivity of future researchers. Future R&D firms respond by hiring more researchers and, thus, further builds the knowledge stock. Both the expansion in the future stock of knowledge and the expansion in the future number of researchers contributes toward increasing the future production of patents, and hence, determine the value of these knowledge spillovers. The elasticity of scale in the R&D production function determines whether the value of these knowledge spillovers expands or contracts as the stock of knowledge grows over time–when the price of new patents is constant over time.

The second set of simulation results is presented in Figure 13.2 for different values of the elasticity of scale in the R&D production function. In these scenarios, the price of new patents is assumed to increase over time due to an increasing price of carbon emissions.

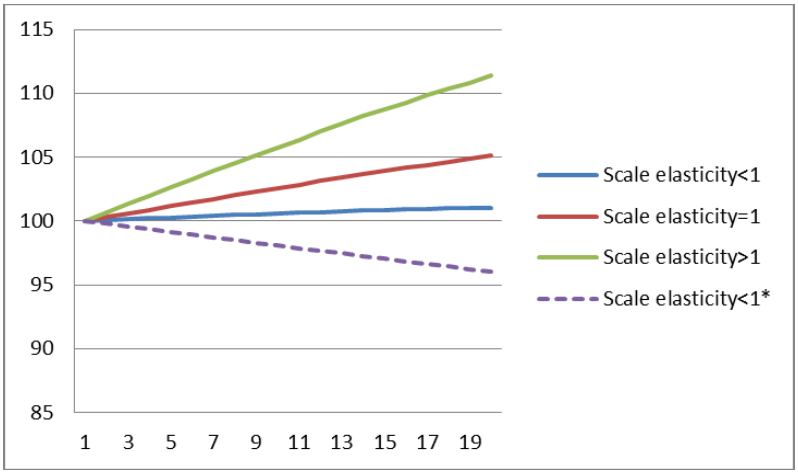

FIGURE 13.2: Optimal subsidy rates with increasing patent price

The initial subsidy rate within each scenario in Figure 13.2 is set equal to 100 to facilitate a comparison of scenarios. Figure 13.2 illustrates that the optimal sub-

sidy rates increase over time when the elasticity of scale equals one, and when the elasticity of scale exceeds one. Further, the optimal of subsidy rates also increase over time in a case where the elasticity of scale is set slightly below one, while they decrease over time in another case where the elasticity of scale is set even lower (indicated by 1*). The intuition is that the impact of R&D subsidies on knowledge spillovers and production of future patents is identical with the case above, where the price of new patents was assumed to be constant over time. The price of new patents is, however, assumed to increase over time in this case. The value of boosting the production of future patents consequently increases over time. This effect contributes toward generating increasing optimal subsidy rates over time. Hence, this explains why the optimal subsidy rates increase over time when the price of new patents increases over time and the elasticity of scale is set equal to or slightly below one.

The simulation model is constructed with a finite horizon to obtain a numerical solution of the government maximization problem. Knowledge spillovers are limited in later periods by this simplification because the value of knowledge spillovers is reduced to zero in the last period of simulation. We, however, conduct a sensitivity test where we show that an increase in the number of simulation periods from 125 to 150 years only has a marginal impact on optimal R&D subsidy rates for the first 20 years of simulation; see Figure 13.3 where optimal subsidy rates are displayed for these two cases. Hence, we only present and interpret model simulations that are based on the first 20 years of simulation to prevent our results being hampered by the finite horizon. Note that the price of patents is constant over time and the elasticity of scale equals one in both these cases.

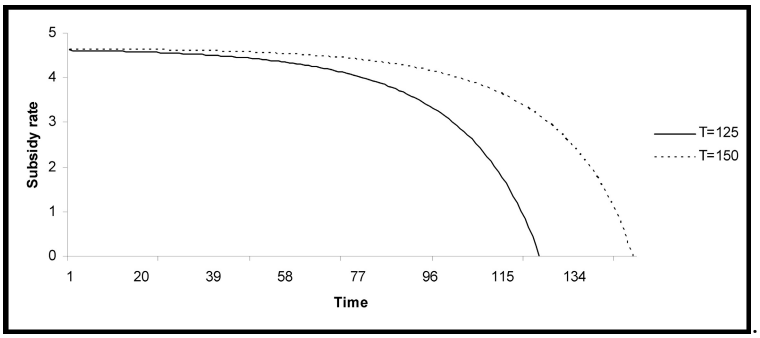

FIGURE 13.3: Optimal subsidy rates with scale elasticity equal to 1

13.3.1 EXTENSIONS

The partial model framework is designed to study the impact of knowledge spillover on the timing of R&D policy. This section extends the model framework to investigate aspects that are omitted in the previous section. First, we investigate implications of introducing a government budget constraint, where the amount allocated to R&D subsidies is restricted. Second, we investigate implications of assuming that old patents become obsolete.

13.3.1.1 Constrained R&D subsidies

Public spending is to a large extent financed by distorting taxes in most countries. Hence, governments are inclined to impose a cost-benefit requirement on public spending that reflects that the marginal cost of public funds exceeds one. We investigate the implications of such a cost-benefit requirement on the timing of R&D policy by imposing a government budget constraint where the amount of public resources allocated to R&D subsidies is restricted. This restriction implies that the government is forced to trade off the benefits of awarding subsidies in one period against the benefits of awarding subsidies in other periods. Model simulations of previous scenarios are conducted with a government budget constraint where the present value of R&D subsidies awarded to R&D firms amounts to 50 percent of the present value of R&D subsidies awarded to firms in previous scenarios without restrictions on public spending. The government budget constraint BC is given by

$$BC = \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} z_t L_t .$$

The simulations with governmental budget constraints show that the timing of optimal R&D subsidy rates is unaffected by the constraints. The constraints, however, lower optimal subsidy rates in all periods. The intuition is that there is an additional cost connected to awarding R&D subsidies to firms, and that this cost is imposed on subsidies in all periods with equal magnitude. Hence, optimal subsidy profiles are virtually unaffected by the constraint.

13.3.1.2 Creative Destruction

Patents and varieties of goods are likely to become obsolete at some point in time as new and improved innovations emerge. Creative destruction is not incorporated into our main model framework. We, however, shed light on the implications of

this aspect by assuming that obsolete varieties are removed from the stock of knowledge that contributes toward generating new patents. To allow for this, we expand our numerical model to include depreciation of knowledge, represented by a constant rate of depreciation. The function for the knowledge stock is updated to

$$A_t = (1-\delta)A_{t-1} + X_t,$$

so that

$$A_t = \sum_{i=0}^{t}(1-\delta)^{t-i} X_i,$$

where δ is the rate of depreciation. Model simulations show that our previous results hold when a modest rate of depreciation is introduced so that the stock of knowledge grows over time. The intuition is that the timing of externalities connected to knowledge spillovers is unaffected, even though the level of externalities is reduced. More drastic rates of depreciation, however, lead to drastic changes in our simulation results. Model simulations show that optimal subsidy rates are constant over time for any elasticity of scale if the production of new patents equals the depreciation of patents, so that the stock of knowledge is constant over time. The intuition is that the government maximization problem becomes identical at the beginning of any period. Thus, the optimal subsidy rate is constant.

Model simulations show that our previous results are completely reversed when a drastic rate of depreciation is introduced so that the stock of knowledge declines over time. The optimal combination of subsidy rates decrease (increase) over time when the elasticity of scale is larger (smaller) than one and the price of patents is constant over time. The intuition is that the stock of knowledge is shrinking. Hence, externalities connected to R&D are shrinking (expanding) with increasing (decreasing) returns scale. Some studies have identified substantial R&D depreciation rates; see Bernstein and Mamuneas (2006) and Mead (2007). A substantial rate of depreciation that leads to a decreasing stock of knowledge does not seem to be empirically relevant as the global warming problem is in an emerging stage.

13.4 CONCLUSION

How governments should engage in policies to spur environmental R&D activity from private firms is an important policy question, since research markets are riddled with inefficiencies. In this paper we explore how one of these inefficiencies–externalities from knowledge spillovers–affects the optimal timing of subsidies for environmental R&D. Model simulations show that optimal R&D subsidy rates

increase (decrease) over time when the elasticity of scale of the R&D production function is larger (smaller) than one–and the price of new patents are constant across time. Model simulations also show that optimal R&D subsidy rates are increasing when the elasticity of scale of the R&D production function equals (or is slightly below) one–and the price of patents increases over time. The majority of the empirical evidence in combination with our simulation results supports the conclusion that subsidies to firms conducting environmentally friendly R&D should increase over time. We also show that a stricter government budget constraint influences the optimal level of R&D subsidies, while the timing issue is unaffected. However, a rate of depreciation on the stock of knowledge may affect the timing of subsidies.

There are some caveats to our conclusion. First, there is no uncertainty in our model. Including a probability of successful innovation would lower the incentives for firms to conduct R&D. Lower R&D activity implies that the social value of increasing the productivity of R&D goes down, and the government would respond by lowering the level of R&D subsidization. If firms get better (worse) at screening R&D projects over time, interpreted as an increase (decrease) in the probability of success, this would increase (decrease) the government's incentives to subsidize over time. We are not aware of any empirical evidence on whether the probability of success increases or decreases over time.

Second, the empirical literature on output elasticities in the R&D production function is not very well developed. Further research is needed to establish significant ranges for the output elasticities. Third, we have only included one type of externality in the research market. Other externalities, e.g. monopoly pricing and research congestion, may also influence the optimal timing of subsidies to environmental R&D. This is a venue for future research.

REFERENCES

Abdih, Y. and Joutz, F. (2005). Relating the knowledge production function to total factor productivity: an endogenous growth puzzle, IMF Working paper 05/74.

Acemoglu, D., Aghion, P., Bursztyn, L., and D. Hemous (2012). The Environment and Directed Technical Change. *American Economic Review* 102, 131–166.

Acemoglu, D., Akcigit, U., Hanley, D., & Kerr, W. (2016). Transition to clean technology, *Journal of Political Economy*, 124(1), 52–104.

Aghion, P., Howitt, P. (1992). A model of economic growth through creative destruction, *Econometrica* 60, 323–351.

Bernstein, J. and Mamuneas, T (2006). R&D depreciation, stocks, user costs and productivity growth for US R&D intensive industries, *Structural Change and Economic Dynamics* 17(2006) 70–98, Elsevier.

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry, *Econometrica*, 81(4), 1347–1393.

Carraro C., Grelagh R., van der Zwaan B.C.C., (2003). Endogenous Technological Change in Economy-Environment Modeling, an Introducton, *Resource and Energy Economics* 25, 1–10.

Dechezleprêtre, Antoine, Ralf Martin, and Myra Mohnen (2017). Knowledge spillovers from clean and dirty technologies: A patent citation analysis. Grantham Research Institute on Climate Change and the Environment Working Paper No. 135.

Fischer, C., & Heutel, G. (2013). Environmental macroeconomics: Environmental policy, business cycles, and directed technical change. *Annu. Rev. Resour. Econ*., 5(1), 197–210.

Gerlagh, R., Kverndokk, S., and Rosendahl, K.E. (2009). Optimal Timing of Climate Change Policy, *Environmental and Resource Economics* 43, 369–390.

Gerlagh, R., S. Kverndokk and K.E. Rosendahl (2014). The optimal time path of clean energy R&D policy when patents have finite lifetime, *Journal of Environmental Economics and Management* 67, 2–19.

Gong, G., Greiner, A., and Semmler, W.(2004). Endogenous Growth: Estimating the Romer Model for the US and Germany, *Oxford Bulletin of Economics and Statistic*s 66, 147–164.

Grossmann, V., Steger, T.M., and Trimborn, T. (2013). Dynamically Optimal R&D Subsidization, *Journal of Economic Dynamics and Control*, 37(3), 516–534.

Greaker, M, Heggedal, T.R., and Rosendahl, K.E. (2017). Environmental Policy and the Direction of Technical Change, forthcoming *Scandinavian Journal of Economics* DOI: 10.1111/sjoe.12254

Griliches, Z. (1995). R&D and Productivity: Econometric Results and Measurement Issues, in Handbook of the Economics of Innovation and Technical Change, P. Stoneman (Ed.), Blackwell, Oxford.

Heggedal, T.R (2015). Knowledge spillovers and R&D subsidies to new, emerging technologies, *Economics of Innovation and New Technology*, 24(7), 710–733.

Heggedal, T.R. and Jacobsen, K. (2011). Timing of innovation policies when carbon emissions are restricted: An applied general equilibrium analysis, *Resource and Energy Economics* 33, 913–937.

Heggedal, T.R., Moen, E.R., and Preugschat, E. (2017). Productivity Spillovers Through Labor Mobility in Search Equilibrium, *Journal of Economic Theory*, 169, 551–602.

Hart, R. (2018). To everything there is a season: Carbon pricing, research subsidies, and the transition to fossil-free energy, forthcoming *Journal of the Association of Environmental and Resource Economist*.

Hoel, M. and Kverndokk, S. (1996). Depletion of Fossil Fuels and the Impacts of Global Warming, *Resource and Energy Economics* 18, 115–136.

IEA 2017. World energy outlook 2017.

Jaffe, A., Newell R. G., Stavins R. N. (2005). A tale of two market failures: Technology and environmental policy, *Ecological Economics* 54, 164–174.

Jones, C. I. (1995). R&D based models of economic growth, *Journal of Political Economy* 193, 759–84.

Jones, C. I. (1999). Growth: With or Without Scale Effects, *American Economic Review* 89, 139–144.

Jones, C.I. and Williams, J.C. (2000). Too Much of a Good Thing? The Economics of Investment in R&D, *Journal of Economic Growth* 5, 65–85.

Klette, T.J., Møen, J., and Griliches, Z. (2000). Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies, *Research Policy* 29, 471–495.

McGlade, C., and Ekins, P. (2015). The geographical distribution of fossil fuels unused when limiting global warming to 2 C. *Nature*, 517(7533), 187–190.

Mead, C. I. (2007). R&D Depreciation Rates the 2007 R&D Satellite Account, Bureau of Economic Analysis/National Science Foundation, R&D Satellite Account Background Paper, November 2007.

Millar, R. J., Fuglestvedt, J. S., Friedlingstein, P., Rogelj, J., Grubb, M. J., Matthews, H. D., and Allen, M. R. (2017). Emission budgets and pathways consistent with limiting warming to 1.5 C. *Nature Geoscience* 10, 741.

Nordhaus, W.D (2008). A Question of Balance: Weighing the Options on Global Warming Policies, Yale University Press, New Haven.

Perez-Sebastian, F. (2007). Public support to innovation and imitation in a non-scale growth model, *Journal of Economic Dynamics & Control* 31, 3791–3821.

Pessoa, A. (2005). ”Ideas” driven growth: the OECD evidence, *Portuguese Economic Journal* 4, 46–67.

Popp, D. and Newell, R. G. (2012). Where does energy R&D come from? Examining crowding out from environmentally-friendly R&D, *Energy Economics*, 34(4), 980–991.

Porter, M.E, and Stern, S. (2000). Measuring the “ideas” production function: Evidence from international patent output, Working paper 7891, National Bureau of Economic Research.

Romer, P. (1990). Endogenous Technological Change, *Journal of Political Economy* 94, 1002–1037.

Samaniego, R.M. (2007). R&D and growth: The missing link?, *Macroeconomic Dynamics* 11, 691–714.

Smulders, S. and Bretschger, L. (2000). Explaining environmental Kuznets curves: how pollution induces policy and new technology, Discussion Paper 95, Tilburg University, Center for Economic Research.

UNFCCC (2015). Adoption of the Paris Agreement FCCC/CP/2015/L.9/Rev. 1; http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf

13.5 APPENDIX A: PARAMETER LIST

The following values were used in all simulations:

A_0	500000
P	1
r	0.07
T	125
X_0^{UE}	12500

UE = Unregulated equilibrium

The tables below show the different parameter values for ϕ and λ that were simulated. The values marked with * are the values used in the Figure 13.1.

Scale elasticity = 1

$$\begin{array}{c} \phi \\ \lambda \end{array} \begin{bmatrix} 0.1 \\ 0.9 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix} \begin{bmatrix} 0.3 \\ 0.7 \end{bmatrix} \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix} \begin{bmatrix} 0.5^* \\ 0.5^* \end{bmatrix} \begin{bmatrix} 0.55 \\ 0.45 \end{bmatrix} \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix} \begin{bmatrix} 0.7 \\ 0.3 \end{bmatrix}$$

Scale elasticity < 1

$$\begin{array}{c} \phi \\ \lambda \end{array} \begin{bmatrix} 0.05-0.09 \\ 0.9 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.25-0.6 \end{bmatrix} \begin{bmatrix} 0.1-0.54 \\ 0.45 \end{bmatrix} \begin{bmatrix} 0.25 \\ 0.1-0.25 \end{bmatrix} \begin{bmatrix} 0.35 \\ 0.35 \end{bmatrix} \begin{bmatrix} 0.4-0.55 \\ 0.1 \end{bmatrix}$$

$$\begin{array}{c} \phi \\ \lambda \end{array} \begin{bmatrix} 0.45 \\ 0.1-0.54 \end{bmatrix} \begin{bmatrix} 0.7 \\ 0.2-0.29 \end{bmatrix} \begin{bmatrix} 0.475^* \\ 0.5^* \end{bmatrix}$$

Scale elasticity > 1

$$\begin{array}{c} \phi \\ \lambda \end{array} \begin{bmatrix} 0.105 \\ 0.9 \end{bmatrix} \begin{bmatrix} 0.45 \\ 0.56-0.59 \end{bmatrix} \begin{bmatrix} 0.56-0.6 \\ 0.45 \end{bmatrix} \begin{bmatrix} 0.525^* \\ 0.5^* \end{bmatrix} \begin{bmatrix} 0.55 \\ 0.55 \end{bmatrix} \begin{bmatrix} 0.7 \\ 0.31-0.39 \end{bmatrix}$$

DOI: 10.18261/9788215031583-2018-15

Chapter 14
Seasoned parliamentarians perform worse than students in a lobbying experiment[1]

LEIF HELLAND, LARS CHR. MONKERUD AND GJERMUND LØYNING

ABSTRACT We present results of a laboratory experiment on costly lobbying, comparing the behavior of elite politicians and students. Our main finding is that members of the Norwegian national assembly deviate more from equilibrium predictions than students. This is in opposition to earlier experimental findings comparing the behavior of students and experienced public relations officers. Our finding is somewhat troubling, given that the underlying model addresses experienced real-world, decision makers. Ours is the first systematic study using members of a national parliament as subjects in a lobbying experiment.

KEYWORDS: Decision-making | Laboratory experiment | Lobbying

14.1 INTRODUCTION

In this chapter, we depart from a stylized game theoretic model of the interaction between a lobbyist and a decision maker. Thus, we are addressing an issue in the interface between business and politics. The core implications from the model are tested in a highly controlled environment: a laboratory experiment. Laboratory experiments have become an important and established tool in business studies over the last couple of decades.[2] In the experiment we tap into a highly unusual

1. Financial assistance from BI Norwegian Business School is greatly appreciated. Thanks to Jan Potters and Frans van Winden for generously sharing material underlying their 2000 experiment. We are indebted to Rolf Aaberge and Alexander Cappelen for constructive comments on an earlier draft.

2. An excellent overview of laboratory experiments in operations management is provided by Donohue et al., (2018). To assess the breadth of the experimental program in current business studies, it is a good idea to consult recent issues of *Management Journal*, the top journal in the field.

pool of subjects, namely seasoned members of the Norwegian National Assembly (the Storting). Their behavior in the lobbying experiment is compared to the behavior of master students at BI Norwegian Business School.

Special interests are an integral part of democratic decision making, and they expend significant resources in their attempts to impact public policy.[3] Frequently, influence is sought by strategically transmitting private information to policy makers; or, in short, by lobbying.[4]

Recent decades has witnessed advances in the understanding of lobbying.[5] A core insight is that lobbying costs can increase the informative content of lobbying messages in equilibrium, and thereby improve democratic decision making.

We investigate experimentally the canonical model of costly lobbying (Potters and van Winden 1992).[6] In the model there is only one lobby. Nature is in one of two possible states, and payoffs are state dependent. The lobby is privately informed about the true state. The lobby moves first. It chooses whether to send a costly signal (to lobby), or not to send a signal at all. The decision maker moves last. She confronts a binary choice; either preserve the status quo, or implement a fixed alternative policy. Independent of the true state of nature, the lobby prefers the alternative policy. In the absence of new information, the decision maker's prior belief favors the status quo. For an interesting range of cost–payoff parameters, a strategically rich signalling game results in which semi-separating and pooling equilibria coexist. Various refinements can be employed in order to select between the equilibria of the model. The stylized environment of the model captures core trade-offs in the interaction between lobbyists and decision-makers.

This model has been subjected to experimental tests (Potters and van Winden 1996; 2000). One such test (Potters and van Winden 2000) compares behavior in two distinct subject groups: students and experienced public affairs officials (pro-

3. Estimates of expenses and activities for special interests in the federal US process are provided by Grossman & Helpman (2001:chapter 1). For a critical interpretation of such numbers, see Ansolabere et al., (2003). For interest group activity in the EU, see the survey of Mahoney (2004) with references.
4. Grossman and Helpman (2001:4–13), and Sloof (1998:18–20).
5. See for instance Grossman and Helpman (2001, part I)I; Persson and Tabellini (2000, ch.7).
6. The model has been extended in various directions. An overview of many extensions, and a thorough analysis of a few more, is provided by Sloof (1998). For a model that departs from slightly different assumptions (two lobbies), but reach qualitatively comparable conclusions for certain parameters, see Austen-Smith and Wright (1992). Austen-Smith and Wright (1996) provide tests of the model on field data from the US Congress.

fessionals).[7] They find that professionals behave more in accordance with the model than students, thereby achieving a higher degree of separation and higher earnings. They conclude that professionals outperform students in playing the lobby game in a controlled environment.

Comparing inexperienced students and subjects with (model) relevant real-life experiences provides an external validity check on laboratory findings (Ball and Cech 1996). The approach is fairly common.[8] Confidence in the empirical content of the underlying model is strengthened (weakened) if relevant professionals outperform (are outperformed by) students.

We subject Potters and van Winden's (1992) costly lobbying model to an additional test of external validity: by comparing the behavior of students and representatives of the Norwegian National Assembly. To the best of our knowledge, ours is the first study that systematically uses elite politicians as subjects in a controlled laboratory experiment on decision-making.[9] Our main finding is that elite politicians behave less in accordance with the costly lobbying model than students do. Elite politicians, having extensive real-life experience with lobbying, achieve lower degrees of separation and lower expected gains than inexperienced students. In our opinion, these results challenge the external validity of the costly lobbying model.

The remainder of the paper is organized as follows. In the next section our design, hypotheses, and procedures are outlined. Thereafter we present our result, and end the paper with a brief conclusion.

---

7. Professionals are described as executives who subscribed to two different conferences on public affairs (one in Amsterdam, the other one in The Hague). In particular, they held positions as public affairs and public relations officers, in the public as well as the private sector.
8. For instance: Cooper (2006) and Cooper et al., (1999) compare managers and students in coordination experiments; Dyer et al., (1989) compare students with experienced contractors in experiments on common value auctions; Dejong and Forsythe (1986) compare businessmen and students in sealed offer experimental markets; Alantas et al., (2006) compare (Indonesian) students and public servants in a corruption experiment; Abbik and Rockenbach (2006) compare professional traders with students in an options-pricing experiment; Wooders (2010) compares students and soccer professionals' ability to play minimax in an experiment on a zero-sum game. Alevy et al., (2007) compare data collected in a field experiment on information cascades using professional traders from the Chicago Board of Trade, with laboratory data using students.
9. Fatas et al., (2007) explore framing effects in a survey experiment including Spanish elite politicians. They find weak evidence of less deviation from the rational-choice model among elite politicians.

14.2 DESIGN, HYPOTHESIS AND PROCEDURES

We start by presenting the parametrized version of the game played by our subjects, the equilibria of this game, and the ensuing behavioral predictions that are tested. Subsequently, we outline our experimental procedures. Our game parameters are identical to the ones used by Potters and van Winden (2000). Apart from some (innocuous) economizing on language in the instructions used, and the employment of a different payoff procedure, we seek to replicate the experimental protocol of Potters and van Winden (2000) faithfully.

14.2.1 THE GAME

The signaling game under study has two players, a sender (S) and a receiver (R). The time line is as follows; first nature draws a black ball with probability 1/3 or a white ball with complementary probability 2/3; the outcome of the draw is observed by S, but not by R; S decides on whether to send a costly signal ($c > 0$) (which may either have the content "the ball is black" or the content "the ball is white"), or to send no signal at all ($c = 0$); R observes the signal, updates his prior probability assessment on the color of the ball, and takes the decision $B1$ or $B2$, finally payoffs are distributed and the game ends.

Implementing the game with neutral language, such as white ball and black ball instead of for instance bad state and good state; sender and receiver instead of lobbyist and politician; and signal instead of lobbying was done in order to minimize the effect of context on behavior.

The parameters of the game are reproduced as a state-decision matrix in Table 14.1 (cost of signaling and *a priori* probability of state in parenthesis). In the table, the payoffs of S are provided first, then the payoffs of R. The strategic tension lies in the fact that R prefers $B2$ only if the state is black ball (the good state), while S always prefers $B2$. Evidently, signals cannot be trusted at face value. In deriving the equilibria of the model, the rules of the game and the information contained in Table 14.1 is assumed to be common knowledge.

TABLE 14.1: Parameters of the experiment.

		***R*'s choice**	
		***B*1**	***B*2**
Low cost	White (2/3)	2, 3	4, 1
(c = 0.5)	Black (1/3)	1, 0	7, 1
High cost	White (2/3)	1.5, 3	3.5, 1
(c = 1.5)	Black (1/3)	1.5, 0	5.5, 1

The game has two equilibria for the parameters considered.[10] The first one is a pooling equilibrium in which *S* sends no message in either state, and *R* always chooses *B*1. Since all sender types send the same signal, it reveals no new information. The prior then favors the decision *B*1. If no receiver chooses *B*2, incurring a cost by sending a signal is futile.

The second equilibrium is separating. In this equilibrium, *S* sends a costly message for sure if the ball is black, and with probability 1/4 if the ball is white. *R* responds to no message by choosing *B*1 for sure. If costs are high, *R* responds to a costly signal by choosing *B*2 with probability 3/4. If costs are low, *R* responds to a costly signal by choosing *B*2 with probability 1/4.

The logic in this equilibrium is as follows. After observing a costly message, *R*'s update (according to Bayes rule) is $Pr(Black\ ball \mid c > 0) = 2/3$. Given the update, *R* is indifferent in her choice. A best reply is then to randomize over her choice so as to make *S* indifferent between sending a costly message or no message at all. If costs are high, this is achieved for a probability of choosing *B*1 equal to 3/4. If costs are low, this is achieved for a probability of choosing *B*1 equal to 1/4. Given this response by *R*, sending a costly message for sure if the ball is black, and with probability 1/4 if the ball is white, is a best reply for *S*. Note that the content of the message does not matter in equilibrium. What matters is whether or not *S* incurs a cost in sending the message.

It has been shown that the separating equilibrium, but not the pooling equilibrium, passes the "universally divine" refinement criteria (for out of equilibrium beliefs), suggested by Jeffrey Banks and Joel Sobel in 1987. In Potters and van Winden's (1992, 2000) opinion, this favours the separating equilibrium as a behavioral prediction.[11]

10. A general treatment of the equilibria in the costly signalling game is provided in Potters and van Winden (1992), and in Sloof (1998).
11. Using this (and other sophisticated) refinement criteria to underpin behavioral predictions is not uncontroversial. Se, for instance, the discussion in Samuelson 1997:5–12.

14.2.2 HYPOTHESES

We identify six testable hypotheses. Departing from the separating equilibrium, the following three follow immediately.

Hypothesis 1 (Signaling): a) the probability of a costly signal is higher following a black ball than a white ball; b) the probability of decision *B*2 is higher following a costly signal than no signal.

Hypothesis 2 (Treatment effect, receivers): a) the probability of decision *B*2 following a costly signal is higher in the high cost treatment than in the low cost treatment); b) the probability of decision *B*2 following no signal is independent of treatment.

Hypothesis 3 (Treatment effect, senders): a) the probability of a costly signal following a black ball is independent of treatment; b) the probability of a costly signal following a white ball is independent of treatment.

The next three hypotheses stipulate that elite politicians are better at playing the lobby equilibrium than students, both in the roles of senders and receivers. Successful candidates in elite politics are selected in highly competitive environments. One trait they are selected on is their ability to effectively signal their types in political campaigns. Such signalling can be modelled in ways that are structurally identical to the lobby game sketched above, albeit with the candidate as sender and the voter as receiver (Sloof 1998:55–8, and chapter 4).

Successful candidates enter elite politics, and become exposed to lobbyists on a regular basis in the role of receivers. At times, the efficient handling of such relationships can be consequential for political survival.

Empirics on the regularity of exposure can be found in a survey of members of the Norwegian Storting in 2001. Representatives were asked: How common is it to be contacted by professional lobbyists? By professional lobbyists in this context is meant hired guns, that is, lobbyists working for a customer on a contract-by-contract basis (as opposed to identifiable special interests that lobby in established corporatist channels). Norwegian institutions do not require professional lobbyists to disclose the identity of customers, and professionals are careful to protect their identity. Thus, the survey question addresses a relationship where precise prior information on preferences is largely unavailable, and signaling ought to be of particular importance.

Responses to the survey question were conditioned on the issue considered most important to the respondents in the current session (presumably an issue of ultimate consequence for political survival), and other issues. Above 20% of the representatives indicated that some or all of the organizations active in the issue

considered most important used professional lobbyists. Close to 90% responded likewise with respect to the other issues category (Gullberg and Helland 2003).

Students, of course, are not selected on their ability to act effectively as senders in signalling games, and do not acquire substantial experience as receivers in signalling games through their studies.

The fourth hypothesis stipulates that the behavior of parliamentarians is closer to equilibrium behavior of senders than is the behavior of students. We use the two measures of signal error suggested by Potters and van Winden (2000) to check on this. The first one is an unweighted measure: $1/2|\sigma_b - S_b| + 1/2|\sigma_w - S_w|$, where $\sigma_b = 1$ is the equilibrium frequency of costly signals after a black ball, $\sigma_w = 1/4$ is ditto after a white ball, and S_b, S_w are actual frequencies of costly signals after a black ball and a white ball, respectively. The second measure uses the prior probability of drawing a black and white ball respectively as weights: $1/3|\sigma_b - S_b| + 2/3|\sigma_w - S_w|$, with notation as in the first measure.

Hypothesis 4 (Gamesmanship senders): politicians have lower signal errors than students.

The fifth hypothesis relates to the level of gamesmanship achieved by politicians in the role they presumably master best; as receivers. Decision errors are defined in the following way $1/2|\rho_0 - \beta_0| + 1/2|\rho_k - \beta_k|$. In the formula, ρ_0 is the equilibrium frequency of $B2$ decisions after observing no-signal; ρ_k is the equilibrium frequency of $B2$ decisions after observing a costly signal, which is conditioned on the cost treatment $k = \{C_L, C_H\}$; β_0 is the actual frequency of $B2$ decisions given no signal; and β_k is the actual frequency of $B2$ decisions given a costly signal. We believe that politicians will make more correct decisions (i.e. be closer to equilibrium) than students.

Hypothesis 5 (Gamesmanship receivers): politicians have lower decision errors than students.

The last hypothesis is more *ad hoc*, and stipulates that politicians—who need a measure of trust in order to get (re)elected—tend to follow the dictums of honesty more closely than students. An implication is that politicians are less likely than students to send costly signals unless a black ball has in fact been drawn. A similar hypothesis is used for the professionals in Potters and van Winden (2000). Their measure of separation is: $(S_b^i - S_w^i)$, $i = (Students, Politicians)$, with S_b and S_w defined as above.

Hypothesis 6 (Separation): politicians achieve a higher degree of separation than students.

14.2.3 EXPERIMENTAL PROCEDURES

In Potters and van Winden (2000), some sessions are conducted as computerized experiments, others as pen and paper experiments. Results are robust to this variation. We replicate with computerized sessions only. Potters and van Winden (2000) programmed their experiment with software no longer in use.[12] Our sessions were programmed in zTree (Fischbacher 2007). We were not able to obtain screenshots from the original computerized sessions. Screenshots were therefore replicated based on information available in written instructions, working papers and final reports.

The written instructions for the computerized sessions in Potters and van Winden (2000) were in Dutch. These were translated to Norwegian with the help of two independent translators. One translated from Dutch to Norwegian, the other compared the translated Norwegian text with the original Dutch text. We had extensive communication with the translators in this process.

Instructions concerning payoffs were re-written to match the payoff structure in our lottery sessions (see below). In addition, some minor adjustment were made in order to economize on language and align the instructions with the typical format used in present day experiments.

Students were recruited from among first-year master students at BI Norwegian Business School via student e-mail and during lecture breaks. None of the recruited students had prior knowledge of the experiment, or experience as subjects in previous economics experiments at the school. No subject was used in more than one session.[13]

To attract parliamentarians for the experiment, a seminar on taxation and redistribution (including lunch) was arranged at the school.[14] The idea was for this seminar to function as a showup fee in kind. Effort was made to pick a day that avoided clashes with essential business in the Storting; to announce the seminar and experiment widely in relevant networks; as well as to secure standbys in the (highly likely) case of last minute drop outs. Forty-seven out of the 169 elected

12. The computerized sessions in Potters and van Winden was programmed in EASEL for the OS 2 operating system of IBM.
13. Subjects used in economics experiments at the school are entered in a historical data base. For an evaluation of the representativeness of student populations used in lab experiments relative to the population at large see Egas and Reidel (2008), Dohmen et al., (2008), and Belot et al., (2010). For an evaluation of the representativeness of students participating in lab experiments relative to students that do not, see Cleave et al., (2011), Falk et al., (2013).
14. Two leading Norwegian experts presented their views on the need for a tax reform: professor in tax legislation at BI Norwegian Business School, Ole Gjems Onstad, and senior researcher Rolf Aaberge at Statistics Norway.

representatives to the Storting responded positively to the invitation. Due to a busy schedule, 12 present and two former parliamentarians eventually participated in the experiment. In addition, six (non-elected) political advisors from the parliamentary party groups were recruited. Only elected representatives played in the role as responders (*R*) in the sessions with politicians.

In Potters and van Winden (2000), performance based monetary payoffs are used in all sessions (adjusted for differences in going wages of students and professionals). In some of our student sessions, we used performance based monetary payoffs, while in other student sessions we used the binary lottery procedure. In all sessions with politicians, we used the binary lottery procedure (see details in Table 14.2).

Why the binary lottery procedure? Norwegian parliamentarians are likely to consider monetary rewards for participation in research illegitimate, and are generally careful to accept monetary payments for services considered part of their duties as elected representatives. For example, most representatives do not accept money for lectures provided at business schools or universities, and many do not accept monetary rewards for articles published in journals or newspapers.

A potentially attractive alternative to monetary payments is the binary lottery procedure (Roth and Malouf, 1979; Berg et al., 1986). In this procedure, subjects earn points, and these points are lottery tickets in a random draw of a fixed price. Since a von Neuman-Morgenstern utility function is linear in the probabilities of outcomes, this payoff procedure renders individual utility linear in points. Thus, the binary lottery procedure controls for variation in risk preferences over subjects. From a purely theoretical point of view, this is the proper way to reward subjects in experiments, if controlling for risk preferences is desirable. On the flip side, results on how the procedure works in practice are mixed (Camerer, 2002:40–41; Berg et al., 2009). Whenever the binary lottery procedure was used in our experiment, the fixed prize consisted of two bottles of good wine. The procedure was accepted by the politicians as a legitimate form of reward.

Students participating in our lottery procedure sessions were paid 100 NOK as a showup fee. In student sessions with performance pay in NOK, there was no showup fee. On average, a session lasted 45 minutes. The exchange rate of experimental currency to NOK in sessions with monetary payoffs produced an expected payoff (in equilibrium) of 200 NOK per subject, which is above the typical optional wage for a student.

The lottery procedure was implemented in the following way. At the end of the session, the sum total of points earned by a subject in that session was divided by a fixed and publicly announced denominator common to all subjects of that ses-

sion, resulting in the assignment of a number between 0 and 1 to each subject. A draw was then made from a random uniform variable in the interval 0 to 1. If the assigned number of a subject was larger (smaller) than the randomly drawn number, this subject would win (not win) a prize consisting of two bottles of good wine. The fixed denominator was calibrated in such a way that half of the subjects in a treatment could be expected to win a prize.

In all student sessions, the experiment was conducted twice. First we ran a session with one cost treatment, then a session with the other cost treatment. The participants were not informed about the second session before the first session had been concluded. Subjects played in the same role in both sessions. Because of time constraints, only one treatment was conducted in each of the two sessions with politicians. Since data from second sessions are likely to be biased (due to dependencies of observations over time), we only use data from first sessions in our analysis.

After a short introduction in a reception room, the participants were asked to draw roles (*S* or *R*) from an envelope, and were subsequently randomly assigned a number indicating their placement in the lab. After seating the subjects, instructions were read aloud (to achieve public knowledge of rules). Subjects where then allowed some time to read through the instructions independently and ask questions. Lastly, subjects where assigned a few simple questions, to make sure they understood the state-payoff matrixes. Subjects (in all sessions) first participated in a non-paying test round (in order to familiarize themselves with the screens), and subsequently played 10 paying rounds of the game.

The same matching protocol as in Potters and van Winden (2000) was employed. In this protocol, no subject is matched with the same subject for two or more consecutive periods, and no subject meets the same subject more than twice during the experiment. All interactions in the experiment preserved the anonymity of the subjects.

TABLE 14.2: Session characteristics (c_L: low cost treatment; c_H: high cost treatment)

Session	N	Subjects	Payoff	Treatment(s)
Pol1	10	Politicians	Lottery	c_L
Pol2	10	Politicians	Lottery	c_H
Stud1	30	Students	Lottery	c_L
Stud2	30	Students	Lottery	c_H
Stud3	10	Students	Money	c_L
Stud4	10	Students	Money	c_H

14.2.4 THE SAMPLE OF POLITICIANS

The 14 (current and former) parliamentarians were on average 47.2 years old (standard deviation 10.9 years). Males and females where equally represented. On average they had served as elected representatives in parliament for 1609 days at the date of the experiment (standard deviation 1093.8 days; maximum of 3769 days; and minimum 889 days).[15] Participating parliamentarians were distributed quite evenly between parties,[16] while members on the finance committee where clearly overrepresented.[17] Of the 12 currently serving parliamentarians participating in the experiment, 67% where reelected into the new Storting in the general election of 2009. The two former parliamentarians used in the experiment both left the Storting after the general election of 2005.

Of the six political advisors used in the experiment, four were males. Their average age was 32.2 years (maximum of 36 and minimum 27 years). The left wing side of the political spectrum was underrepresented in the group of advisors,[18] while advisors to the parties' finance fractions in parliament were overrepresented.[19]

Clearly the sample is not representative of the Storting at large. For the purpose of this study, however, we believe the distribution of party affiliation, age, gender and committee assignments is not likely to have a significant impact on the result, the reason being that the experiment was framed in neutral language without references to gender, age, party ideology or committee tasks.

14.3 RESULTS

We present the results in two sections. First, we assess behavioral patterns within the lottery payoff procedure. That is, we test for certain predictions following from the model in terms of signalling patterns and treatment effects (hypotheses 1

15. Though the participants must be said to have substantial experience as MPs, it is, at the end of the day, a matter of taste whether one feels that their experience merits the label seasoned or not.
16. Progress Party (Fremskrittspartiet) 2; Conservative Party (Høyre) 3; Christian Democratic Party (Kristelig Folkeparti) 3; Liberal Party (Venstre) 2; Labor Party (Arbeiderpartiet) 2; and Socialist Left party (Sosialistisk Venstreparti) 2.
17. Finance committee 6; Justice committee 1; Labour and Social services committee 1; Church, Education and Research committee 1; Energy and Environment committee 2; Industry committee 1; Foreign Affairs committee 1; Local administration committee 1.
18. Conservative party ("Høyre") 1; Christian Democratic party ("Kristelig Folkeparti") 4; and Center party ("Senterpartiet") 1.
19. Advisors to the finance fraction of a party in parliament 4; Advisors to the local administration fraction of a party in parliament 1; Advisor to the prime minsters' office 1.

through 3) and we check for subject pool effects (hypotheses 4 through 6) using only the sessions in which the wine lottery payoff structure was applied (Pol1,2 and Stud1,2). In particular, this allows for a controlled evaluation of subject pool effects in terms of the expectation that politicians are better decision makers than students. In a subsequent section, we briefly compare results from the lottery and money sessions in order to evaluate the impact of payoff structure on observed patterns. To this end, we evaluate and compare results using only the different sessions among the student population (Stud1,2,3,4).

Before presenting the results, we note the following points pertaining to some quite important analytic choices. First, we use non-parametric tests with $p < 0.10$ (one-tailed) as the threshold for significance: Wilcoxon's signed rank test (WSR) for related samples and the Mann-Whitney U-test (MWU) for independent samples.[20] Second, what counts as an observation? According to Potters and van Winden (2000:509), "[t]here are three possibilities: each play of the game, each individual subject or each session". Potters and van Winden themselves end up using sessions as observations, of which there were 15 in their design. In our study we have 4 comparable sessions.[21] However, only analyzing at the session level risks neglecting differences that are significant at the subject level. While the former strategy is the more conservative, we choose to present patterns at the subject level, noting that analysis at the session level produces no significant results for any of the tests presented below, save for tests of hypotheses 1a and 1b (to which we return shortly).

14.3.1 SIGNALLING, TREATMENT AND SUBJECT POOL EFFECTS

Table 14.3 presents the average proportion of costly signals (as averaged over subject means), conditioned on cost treatment (c_L, c_H), and contingent on whether the draw was a white ball (S_w) or a black ball (S_b).[22] Standard deviations over subject means are presented in parentheses. For convenience, the lower rows in these

20. The WSR test is a non-parametric test for within-subject differences, while the MWU test is a non-parametric test for between-subject differences. A good exposition is provided by Bhattacharyya and Johnson (1977, chapter 15).
21. Potters and van Winden (2000:note 11) are able to demonstrate significant differences between sessions with the same experimental design, indicating dependencies within sessions. Obviously, we are not able to (statstically) test for such differences in our material.
22. As for the content of messages; politicians lie somewhat less than students (13% vs. 20% average lie rate over subjects), and refrain from sending a signal somewhat more often (47% vs. 40% for students), but none of these differences are significant (p=0.29 and p=0.52 respectively).

tables also presents the equivalent averages for the student-money sessions (Stud3,4).[23]

TABLE 14.3: Costly signals by treatment (c_L, c_H), contingent on draw of a white ball (S_w) or a black ball (S_b). Average subject level proportions (standard deviations)

Treatment:	c_L		c_H		Average	
	S_w	S_b	S_w	S_b	S_w	S_b
Politicians	.66 (.26)	.93 (.15)	.20 (.25)	.44 (.33)	.43 (.34)	.69 (.35)
Students Lottery	.39 (.29)	.85 (.17)	.49 (.37)	.85 (.19)	.44 (.33)	.85 (.18)
Average	.46 (.30)	.87 (.16)	.42 (.36)	.75 (.29)	.44 (.33)	.81 (.24)
[Students Money	.63 (.37)	.93 (.15)	.45 (.19)	.70 (.45)	.54 (.29)	.82 (.34)]

In a similar fashion, Table 14.4 displays average proportions of B2 decisions contingent on whether the decision maker is playing in the high or low cost treatment and whether he or she has received a costly signal (S) or not (0).

TABLE 14.4: $B2$ decisions by treatment (c_L, c_H), contingent on no signal (β_0) or costly signal (β_S). Average subject level proportions (standard deviations)

Treatment:	c_L		c_H		Average	
	β_0	β_S	β_0	β_S	β_0	β_S
Politicians	.37 (.42)	.33 (.19)	.27 (.12)	.37 (.44)	.33 (.29)	.35 (.32)
Students Lottery	.12 (.20)	.51 (.30)	.16 (.22)	.36 (.24)	.14 (.14)	.43 (.28)
Average	.18 (.28)	.46 (.29)	.19 (.21)	.37 (.29)	.18 (.24)	.41 (.29)
[Students Money	.00 (.00)	.43 (.22)	.15 (.22)	.68 (.24)	.08 (.17)	.55 (.26)]

Since our focus is on predicted differences between student and politician subjects pools, we concentrate our discussion of results on hypotheses 4–6. First, however, we summarize results for general model predictions as laid out in hypotheses 1a–3b.

First of all, costly signalling is significantly more frequent after a black ball has been drawn (hypothesis 1a). As can be seen from Table 14.3 the overall difference in rela-

23. The next section deals with payoff structure effects.

tive frequency is $\Delta = 0.33$, which is highly significant in a WSR test ($p = 0.000$).[24] Moreover, this result holds in both the student and politician subject pools. The relative frequency of *B*2 decisions is also significantly higher after a costly signal has been sent (hypothesis 1b), with an overall $\Delta = 0.23$ ($p = 0.00$). However, this result is driven entirely by student subjects and is clearly not attributable to politicians' behavior, where the difference of only $\Delta = 0.02$ is not near significant ($p = 0.47$).

As can be seen from Table 14.4, there is little support for hypothesis 2a, i.e. the proposition that *B*2 decisions following a costly signal is more frequent in the high cost treatment. The overall difference in relative frequencies, $\Delta = -0.10$, is in fact the reverse of the expectation. This anomaly is even marginally significant among students ($\Delta = -0.15$, $p = 0.09$), while the small positive difference ($\Delta = 0.05$) for politicians is insignificant ($p = 0.50$).

As for the remaining hypotheses, overall results are in line with expectations, since neither *B*2 decisions following a no-message signal (hypothesis 2b), nor costly signals following the draw of a black ball (hypothesis 3a) or a white ball (hypothesis 3b) are significantly more frequent in either the high cost or low cost treatment ($\Delta = -0.01$, $\Delta = -0.13$, $\Delta = -0.04$ and $p = 0.26$, $p = 0.11$, $p = 0.26$ respectively). However, whereas the results for the latter two hypotheses hold in the student population, there are significant anomalies in politicians' behavior. Here, the differences in relative frequencies of costly signalling ($\Delta = -0.49$ and $\Delta = -0.46$ respectively) are significant ($p = 0.02$ and $p = 0.03$ respectively). We return to these general patterns and anomalies in the the two subject pools below. For now we move on to a more detailed discussion of behavior amongst politicians and student populations (hypotheses 4–6).

Hypothesis 4 (gamesmanship, senders). We apply the same measure for prediction errors (for players in the role of senders) as is used in Potters and van Winden (2000:511). It turns out that both the unweighted and the weighted versions of this measure of out-of-equlibrium behavior is substantially larger for politicians (0.31 and 0.30 respectively) than for students (0.21 and 0.23 respectively). Moreover, the differences are statistically significant in MWU tests ($p = 0.05$ for the unweighted difference and $p = 0.07$ for the weighted difference). This finding is in opposition to expectations. One would expect politicians to perform better than students. At the very least they should have some knowledge of the doings of "[p]ublic affairs managers [who] are professionally skilled to transmit information and to influence the beliefs and behavior of policy makers" (*ibid*. 505), and some experience in having played a structurally similar signalling game as senders in their electoral campaigns.

24. Consulting Table 14.3, the average difference of 0.33 in costly signaling (c_H) between subjects drawing a black and a white ball is $\Delta = \pi(c_H|S_b) - \pi(c_H|S_w) = 0.75 - 0.42 = 0.33$ (in which π stands for proportion).

Hypothesis 5 (gamesmanship, receivers). The perhaps most interesting difference between the two subject pools concerns behavior in the role of receivers, i.e. the role to which politicians should arguably be especially accustomed. Applying our measure of out-of-equlibrium behavior ($1/2\left|\rho_0 - \beta_0\right| + 1/2\left|\rho_{S;C_k} - \beta_S\right|$) reveals that politicians, even in the role of receivers, have larger decision errors (0.33) than students (0.25). However, this difference is not statistically significant ($p = 0.13$). Nevertheless, the data speak against the general hypothesis that professional experts (here: politicians in the role of receivers) do (significantly) better than laymen.

Hypothesis 6 (separation). We use an equivalent test to the one in Potters and van Winden (2000:505–506). Specifically, the measure in question ($S_b - S_w$) is meant to tap the ability of achieving full separation ($S_b = 1$– and $S_w = 0$), a result that is off-equilibrium ($S_b = 1$ and $S_w = 1/4$) but that will nevertheless leave subjects with higher earnings. Since politicians are presumably more experienced in lobbying situations, we expect them to achieve a higher degree of separation than students.[25] The results, however, go in the opposite direction. While politicians on average achieve a separation measure of 0.26, students achieve a separation measure of 0.41. Moreover, this difference of $\Delta = -0.15$ is significant ($p = 0.09$).

Finally, one may wonder if there are signs of convergence on (perhaps higher) levels of separation over the course of the ten rounds. Since we have only two independent draws per period per subject pool, a same-color draw in the two renders the separation measure undefined. A complete round-by-round mapping of separation is therefore not possible. However, we may compare separation in the first five rounds with that for the last five rounds. Figure 14.1 displays the pattern for the two subject pools. As can be seen from the figure, separation seems to be reasonably stable between early and late parts of the experiment[26], with levels for students consistently higher than that for politicians.

25. In line with Potters and van Winden (2000:505), we may argue that behavior to this end in particular implies that senders should avoid sending deceitful messages when the color is white. Also, one may argue that the existence of professional rules of conduct such as a never cheat or misinform rule is conducive in that respect (*ibid.*). Again, we may argue that politicians' experience with real-life lobbying situations may also result in behavioral influence from what originates more specifically from calculations on the part of lobbyists. Or one may assume that strong, but nevertheless general norms of honesty harbored by politicians themselves also influences behavior in situations that are not defining features of the policy-maker role.
26. There are no significant differences between separation measures in early and late periods, neither for students nor for politicians. This is similar to patterns found in Potters and van Winden's (2000:512) study. The authors find no particular pattern of convergence, but they do, conversely, find that professionals (lobbyists) consistently achieve higher separation than non-professionals (students).

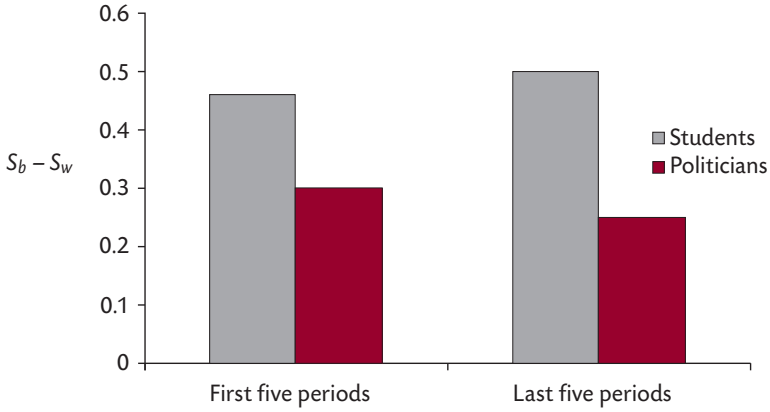

FIGURE 14.1 Separation of students and politicians

14.3.2 PAYOFF STRUCTURE EFFECTS

We now look specifically at differences in behavioral patterns between student subjects playing under the lottery and money payoff structures, respectively. We focus on tendencies for equilibrium (or sensible off-equilibrium) behavior in the two types of games. In other words we parallel the above tests for hypotheses 4 through 6, but now with payoff structure substituting for subject pool categories as the group effect.

First, however, we note that play in the experimental rounds with payment in the form of money (and with students as players) is by and large in accordance with the expectations of hypotheses 1 through 3. With the relevant differences (Δ) calculated from terms in the lower rows of Tables 14.3 and 14.4, the appropriate statistical tests (WSR or MWU tests) reveal that costly signals are sent more frequently after a draw of a black ball ($\Delta = 0.28$, $p = 0.07$, see hypothesis 1a); $B2$ decisions are more frequent in the face of a costly signal ($\Delta = 0.48$, $p = 0.00$, see hypothesis 1b); $B2$ decisions given a costly signal are more frequent in the high cost treatment ($\Delta = 0.26$, $p = 0.10$, see hypothesis 2a); $B2$ decisions in situations of no signal are somewhat more frequent in the high cost treatment ($\Delta = 0.15$, $p = 0.09$, see hypothesis 2b). The last finding is in breach of expectations; the relative frequency of costly signals after a draw of a black ball is independent of whether the game is played in the high or low cost treatment ($\Delta = -23$, $p = 0.22$, see hypothesis 3a); the same goes for the sending of costly signals in the face of a white draw dependent on treatment ($\Delta = -0.18$, $p = 0.34$, see hypothesis 3b).

Turning finally to differences in behavior between the money and the lottery sessions, it turns out that sender gamesmanship is not much different from the one

version to the next. Whereas students in the money game have decision error measures of 0.25 (unweighted measure) and 0.27 (weighted measure), students in the lottery game can show for quite similar measures of 0.21 (unweighted) and 0.23 (weighted) (as reported earlier). Moreover, the differences are not statistically significant ($p = 0.43$ for unweighted measures; $p = 0.40$ for weighted measures). It seems, however, that receiver behavior may be influenced by the payoff structure of the game. While students in the money game, on average, have receiver decision errors of 0.14, students in the lottery game display a somewhat higher error rate of 0.25. Moreover, the difference between the two is statistically significant ($p = 0.04$). Lastly, it turns out that levels of separation do not seem to vary between the two types of games, with students in the money games achieving a slightly inferior separation measure of 0.28 compared to the reported 0.41 for students in the lottery games (the difference not statistically significant at $p = 0.29$).

14.3.3 SUMMING UP THE RESULTS

Statistical tests show that politicians behave less in accordance with predictions from the model than do students. The behavior of the former falls short of expectations on four (hypotheses 1b, 2a, 3a, 3b) out of six counts (hypotheses 1a through 3b). Students perform more in accordance with model expectations, failing to conform to one expectation (hypothesis 2a). Moreover, in terms of behavior politicians perform significantly worse than students in the role of senders and, importantly, do no better than students in the role of receivers. Finally, students achieve significantly higher rates of separation than politicians. We stress that these results pertain to patterns in games where the lottery payoff scheme is applied. However, in supplementary tests of effects from the applied payoff structure, it is found that students by and large conform to model expectations, regardless of payoff scheme.

14.4 CONCLUSION

We have replicated a costly signaling game, comparing the behavior of students and elite politicians. Both groups deviate from equilibrium predictions. However, elite politicians are substantially more off-mark than students. We cannot entirely rule out that our choice of payoff procedure is partially responsible for the results obtained, though the robustness of student behavior to payoff procedure indicates that the lottery procedure might not be culpable for the results. Caution has to be taken due to the low number of observations in our experiment. This said, the main

pattern of more equilibrium deviations by elite politicians seems fairly robust. Most surprisingly, perhaps: elite politicians are no better than students in the role of receivers (interpreting and acting on lobby signals).

In our opinion, and with the appropriate methodological qualifications, our results raise questions pertaining to the external validity of the costly signaling model.

Why do elite politicians deviate more from equilibrium than students? There may be a number of reasons for this. The experiment was conducted without contextualizing it as a lobby problem. A speculation is that a less abstract context could have primed the experience of the politicians more effectively.

Furthermore, any one–or any combination–of the highly stylized assumptions underpinning the lobby model may constitute a poor approximation to the kind of lobby relations experienced by elite politicians in their daily dealings. For instance, real-life interactions are usually repeated under an open horizon, making reputations salient, and expanding the set of equilibria. Also, politicians in real life situations typically encounter more than a single signal. Multiple signals are usually more informative than single signals, and should induce more honesty. If such experiences prime the lab behavior of politicians, it may produce excessive trust (compared to the equilibrium of the single signal model). Our experiment, however, was not designed to isolate reputational concerns and multiple signals.

REFERENCES

Abbink, Klaus and Bettina Rockenbach (2006): Option Pricing by Students and Professional Traders: A Behavioral Investigation. *Managerial and Decision Economics* 27:497–510.

Alevy, Jonathan, Michael Haigh and John List (2007): Information Cascades: Evidence from a Field Experiment with Financial Market Professionals. *The Journal of Finance* LXII(1):151–180.

Ansolabehere, Stephen, John de Figuerido and James Snyder Jr. (2003): Why Is There so Little Money in U.S. Politics. *The Journal of Economic Perspectives* 17(1):105–130.

Atlas, Vivi, Lisa Cameron, Ananish Chaudhuri, Nisvan Erkal and Lata Gangadharan (2006): *Subject Pool Effects in a Corruption Experiment: A Comparison of Indonesian Public Servants and Indonesian Students*. Research Paper No.975: Department of Economics/University of Melbourne.

Austen-Smith, David and John Wright (1992): Competitive Lobbying for a Legislators Vote. *Social Choice and Welfare* 9:229–257.

Austen-Smith, David and JohnWright (1996): Theory and Evidence for Counteractive Lobbying. *American Journal of Political Science* 40(2):543–564.

Ball, Sheryl and Paula-Ann Cech (1996): Subject Pool Choice and Treatment Effects in Economic. Laboratory Research. In: R. Mark Issac (ed.) *Research in Experimental Economics, Vol. 6*. Creenwich: JAI-Press.

Banks, Jeffrey and Joel Sobel (1987): Equilibrium selection in signaling games. *Econometrica* 55:647–661.

Bhattacharyya, Gouri and Richard Johnson (1977): *Statistical Condepts and Methods*. New York: John Wiley and Sons.

Belot, Michèle, Raymond Duch and Luis Miller (2010): *Who should be called to the lab? A comprehensive comparison of students and non-students in classic experimental games*. Discussion Papers No. 2010:001. University of Oxford: Nuffield College.

Berg, Joyce, Lane Daley, John Dickhaut and John O'Brian (1986): Controlling preferences for gambles on units of experimental exchange. *Quarterly Journal of Economics* 101:281–306.

Berg, Joyce, Thomas Rietz and John Dickhaut (2009): On the performance of the lottery procedure for controlling risk preferences. In: Charles Plott and Vernon Smith (eds.) *Handbook of Experimental Economic Results: Volume 1*. Amsterdam: North-Holland.

Camerer, Colin (2002): *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.

Cleave, Blair L., Nikos Nikiforakis and Robert Slonim (2013): Is there selection bias in laboratory experiments? The case of social and risk preferences. *Experimental Economics* 16(3):372–382.

Cooper, David (2006): Are Experienced Managers Experts at Overcoming Coordination Failure? *The B.E. Journal of Economic Analysis and Policy*, Berkeley Electronic Press, 0(2).

Cooper, David, John Kagel, Wei Lo and Qing Liang Gu (1999): Gaming Against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers. *American Economic Review* 89(4):781–804.

Dejong, Douglas and Robert Forsythe (1988): A note on the use of businessmen as subjects in sealed offer markets. *Journal of Economic Behavior and Organization* 9:87–100.

Dohmen, Thomas , Armin Falk, David Huffmann and Uwe Sunde (2008): Representative trust and reciprocity: Prevalence and determinants. *Economic Inquiry* 46(1):84–90.

Donohue, Karen, Elena Katok and Stephen Leider (2018): *The Handbook of Behavioral Operations*. Hoboken, New Jersey: John Wiely and Sons.

Dyer, Douglas, John Kagel and Dan Levin (1989): A Comparison of Naive and Experienced Bidders in Common Value Offer Auctions: A Laboratory Analysis. *The Economic Journal* 99(394):108–15.

Egas, Martijn and Arno Reidel (2008): The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society of London B: Biological Sciences* 275(1637):871–878.

Fatas, E., T. Neugebauer and P. Tamborero (2007): How politicians make decisions: a political choice experiment. *Journal of Economics* 92(2): 167–196.

Falk, Armin, Stephan Meier and Christian Zehnder (2013): Do lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of the European Economic Association* 11(4):839–852.

Gullberg, Anne Therese and Leif Helland (2003): Profesjonell lobbyisme. Norske eliters bruk og holdninger ("Professional lobbyism. Use and attitudes among Norwegian elites"). Oslo: Makt- og demokratiutredningen. Rapport 70/03.

Løyning, Gjermund (2008): *Lobbyism, a waste of time? Parliamentarians and students in a lobbying experiment*. MSc Thesis (Political Economy). Oslo: Norwegian School of Management.

Mahoney, Christine (2004): The power of institutions: State and interest group activity in the European Union. *European Union Politics* 5: 441–466.

Potters, Jan and Frans van Winden (2000): Professionals and students in a lobbying experiment. Professional rules of conduct and subject surrogacy. *Journal of Economic Behavior and Organization* 43(4):499–522.

Potters, Jand and Frans van Winden (1996): Comparative statics of a signaling game. An experimental study. *International Journal of Game Theory* 25:329–354.

Potters, Jan and Frans van Winden (1992): Lobbying and asymmetric information. *Public Choice* 74(3):269–292.

Roth, Alvin and Michael Malouf (1979): Game-theoretic models and the role of information in bargaining. *Psychological Review* 86:574–894.

Samuelson, Larry (1997): *Evolutionary Games and Equilibrium Selection*. Cambridge Mass.: The MIT-Press.

Sloof, Randolph (1998): *Game Theoretic Models of the Political Influence of Interest Groups*. Dordrecht: Kluwer Academic Publishers.

Wooders, John (2010): Does Experience teach? Professionals and Minimax Play in the Lab. *Econometrica* 78(3):1143–1154.

DOI: 10.18261/9788215031583-2018-16

Chapter 15
The Past, Present and Future of Service Marketing:
From Understanding Quality to Understanding Customers

ANDERS GUSTAFSSON AND LINE LERVIK-OLSEN

ABSTRACT Service marketing emerged in response to the shortfalls of product marketing. Although earlier traces exist, it gained traction in the mid 70s. The field's evolution can be divided into phases in which critical incidents can be identified that have led the service field in new directions. Central to the discipline is the service encounter. Research referred to here consists of ways of understanding customer experiences with the service encounter and consequences of the encounter. We summarize various methods or approaches that have been and are applied to understand the service encounter.

KEYWORDS: service encounter | customer experience | customer journey | Norwegian Customer Satisfaction Barometer

15.1 BACKGROUND

The service sector is likely to be the most important sector in any developed economy. In developed countries the service sector now generates more than 70% of a country's GDP (Ostrom et al., 2010). The importance of the service economy is still increasing; the largest growth in the number of firms is in the service sector, which in turn means that the number of employees is also increasing in the sector. New, more service-oriented business models are being implemented, seen in manufacturing companies that are turning into service suppliers. For instance, IBM has totally shifted its business such that it no longer produces goods, but produces and delivers services only. We also see a rise in the collaborative economy and

companies such as Airbnb, Uber, and Craigslist are shifting the competitive rules in service industries (Benoit, Baker, Bolton, Gruber, & Kandampully, 2017). Furthermore, business models are shifting in the retail industry, where companies such as AliExpress and Amazon are forcing changes in competition. To survive, companies must be able to deliver good service with a perception of high quality that is adapted to customers' needs. But this is just a necessary condition; to be competitive, they must go even further. Companies must understand and enhance customer engagement and experience in all parts of any service encounter. Consequently, there is no way around the fact that the service sector is increasing in importance and will continue to be relevant. To be relevant, researchers and organizations must develop new and better tools built to get a deeper understanding of the customer perspective. Companies must balance how much their customers want an organization to know about them with what they should find out. It is a fact that, with still emerging technology, organizations can acquire very deep knowledge about their customers.

Service researchers typically build on Shostack's (1985, p. 243) definition of a service encounter as "a period of time during which a consumer directly interacts with a service." It is during these encounters that customers form their perceptions of service quality. As customers experience multiple encounters, not only directly, but also indirectly through commercials or social media, various encounters may result in loyalty and relationship formations. The focus in a service encounter is usually on the core encounter, but what happens prior to and after the core will also influence the customers' perception of the core service encounter and the overall relationship with an organization (Voorhees et al., 2017).

Service researchers have focused on trying to measure and understand the nuances in service encounters. Parasuraman, Zeithaml, and Berry (1988) developed their SERVQUAL model as an effort to understand the transaction-specific perception of service quality in a service encounter. This was followed by Fornell, Johnson, Anderson, Cha, and Bryant (1996), who recognized that relationships are formed over a series of transactions, a phenomenon known as the cumulative perspective of customer satisfaction. The next step is a realization that not all relationships are created equal and that companies must study portfolios of relationships (Johnson & Selnes, 2004) and understand where in a relationship formation a customer is (e.g., acquaintances, friends, and partners). These ideas were the start of a development to understand how to invest in customers to maximize customer lifetime value, that is, the accumulated cash flow a customer accrues during his or her lifetime (Kumar & Pansari, 2016). Companies need all of these in order to be profitable in the long term. The state-of-the-art thinking of service encoun-

ters from a cumulative perspective is to understand customer experiences over a customer journey across different channels or touchpoints (e.g., online, in-store, and customer service). These customer experiences can cut across a multitude of service providers.

The purpose of this chapter is to give an overview of the above-described development with a focus on understanding the service encounter. We will go back to the start of the field, which is usually stated to be 1977 with a seminal paper by Shostack (1977), and we end by giving one perspective on where the field may be going in the future.

15.2 EARLY STAGES OF RESEARCH IN SERVICE—THE IMPORTANCE OF SERVICE QUALITY

Even if there are earlier traces, Shostack's 1977 article is generally stated as somewhat of a starting point of service as a field of research. She stated that product marketing fell short when marketing services and that "new concepts are necessary if service marketing is to succeed" (Shostack, 1977, p. 73). It was around the same time as the quality movement had started to gain momentum, with important figures such as Crosby, Deming, and Juran. In this school of thought, the quality of products was defined in terms of consistency and low variation in the production. This quality management perspective received extensive attention and enthusiasm, some of which was carried over to service research. For this quality perspective to be properly implemented, service as a phenomenon had to be understood, which was followed by ways to measure and analyze the customer's perspective of a service encounter.

The initial focus in service research was on defining major concepts of what the differences are between services and products. Zeithaml, Parasuraman, and Berry (1985) summarized the discussion on the nature of services to consist of four characteristics, services are intangible, heterogeneous, inseparable, and perishable (the so-called IHIP) and that the focus in service was on the process components rather than on the final product or the outcome of a process (Grönroos, 1998). A joke sometimes told to better describe what a service was, was that a service, as opposed to a tangible good, did not hurt if it was dropped on your foot. Another very important and highly cited concept was the gap model (Parasuraman, Zeithaml, & Berry, 1985). The gap model identifies five gaps—"these gaps can be major hurdles in attempting to deliver a service which consumers would perceive as being of high quality" (Parasuraman et al., 1985, p. 44). The natural conclusion from these concepts was that service quality is seen as something different from

product quality and it must be measured and managed differently. For instance, quality attributes can be categorized as being search, experience, and credence attributes (Zeithaml, 1981). For products, search attributes (attributes that can be determined prior to purchase) are usually more important, while for service, experience and credence attributes are more important. It is from these perspectives that the development of SERVQUAL should be seen. SERVQUAL is designed to measure and understand the service quality in *one* service encounter; in other words, SERVQUAL takes a transaction-specific view because it focuses one transaction. In the SERVQUAL model, service quality is said to be constituted of five dimensions (Parasuraman et al., 1988): *tangibles* (physical facilities, equipment, and appearance of personnel), *reliability* (ability to perform the promised service dependably and accurately), *responsiveness* (willingness to help customers and provide prompt service), *assurance* (knowledge and courtesy of employees and their ability to inspire trust and confidence), and *empathy* (caring, individualized attention the firm provides its customers). The general idea in SERVQUAL is to focus on the dimension that has the largest gap between expectation and performance and close this gap. The SERVQUAL has been very influential and heavily cited, and it does capture the low-hanging fruit when it comes to quality problems. It has, however, been criticized because there is no way of knowing that the identified gaps actually will affect customer satisfaction. In other words, even if there is a large gap between expectation and performance, there is no way of knowing that closing this gap will influence customer satisfaction, because we do not estimate the statistical effect from the gap on customer satisfaction. Also, since the service sector today covers more than 70% of everything that is produced in a country, it includes multiple contexts with multiple understandings of what constitutes quality. In light of this development, SERVQUAL having predefined dimensions and questions might not be sufficient to diagnose details of how to improve business. This paved the way for the next phase of the evolution; instead of focusing entirely on understanding quality, firms started focusing on what drives customer satisfaction and loyalty.

15.3 FROM SERVICE QUALITY TO CUSTOMER SATISFACTION AND LOYALTY

The logic behind the next phase of service research is that service quality is not the only aspect that is important for a company to gain market share. The focus should be on how to change or improve attributes in order to make their customers even more satisfied and in the process make existing customers more loyal and

attract new customers (Anderson, Fornell, & Lehmann, 1994). Also, no company has endless resources and will therefore have to carefully consider where an investment in quality should be made. Resource limitation makes improved quality an optimization process rather than a maximization process; in other words, where should a company invest to get the largest return on investment (Rust, Moorman, & Dickson, 2002). As previously noted, the original interest in marketing and consumer research was on transaction-specific satisfaction, or a customer's experience with a product episode or service encounter. Although the transaction-specific approach had its merits, it does not perform well when predicting subsequent consumer behaviors and economic performance of firms (Fornell et al., 1996; Johnson, Anderson, & Fornell, 1995). In response to the low predictive ability of the transaction-specific approach, researchers started to focus more on what is called cumulative satisfaction (Johnson et al., 1995). The cumulative approach defines satisfaction as a customer's overall experience to date with a product or a service provider; this includes the experience of all service encounters (Johnson & Fornell, 1991). At this stage we would like to point out that we cannot forget about the transaction; it is still important, but organizations need more tools to understand the holistic customer perspective. One of the most well-known approaches in this research is the national customer satisfaction models. The Swedish Customer Satisfaction Barometer (SCSB) model contains two primary drivers of customer satisfaction: expectations about how well the firm would perform when delivering quality and an assessment of how well the firm actually performed (Fornell, 1992). The model contained two consequences of customer satisfaction: customer complaints and customer loyalty. Later, perceived value was added as an antecedent of satisfaction.

The SCSB model is likely to be one of the most well-known models for measuring the causes and consequences of customer satisfaction. Over the years, the SCSB has been used an indicator of various important performance metrics such as market share (Rego, Morgan, & Fornell, 2013), stock market (Fornell, Mithas, Morgeson III, & Krishnan, 2006), and profitability (Anderson, Fornell, & Lehmann, 1994). The national customer satisfaction indexes are useful for the purpose of comparing an organization's performance across industries to get a sense of how well it performs in its own industry and compared to companies in other industries. We will use the Norwegian Customer Satisfaction Barometer (NCSB) to exemplify this comparison. The same approach can, however, be used to measure and manage any organization's more detailed understanding of the customer perspective (Johnson & Gustafsson, 2000). One of the most important aspects therefore is to create a good lens on the customers' perception of the benefits an

organization delivers to its customers. These benefits are not measurable using single indicators but instead must be measured using a latent variable. For instance, easy access as a latent variable may be measured by ease of parking, opening hours, and ease of finding. The benefits will lead to attitudes or a perception of satisfaction which in turn leads to a behavior.

15.3.1 THE NORWEGIAN CUSTOMER SATISFACTION BAROMETER

A change from focusing on service quality to focusing on customer satisfaction as the crucial variable of interest took place throughout the 1990s and in the beginning of the new millennium. It all started in Sweden. Claes Fornell, professor at the University of Michigan, first launched the SCSB already in 1989 (Fornell, 1992). This model served as the prototype when developing the American Customer Satisfaction Index (ACSI) introduced in 1994 and the NCSB introduced in 1996 (Fornell, 1992). In the early days, the NCSB reported results from more than 42 companies across 12 industries; today NCSB reports results from 169 companies across 30 industries annually. NCSB has gained significant influence over the years and represents an important performance metric for companies as well as a benchmark toward competitors. The best performers in each industry, that is, the companies with the highest customer satisfaction scores, typically use their achievements for marketing purposes. The very first NCSB model was identical to the ACSI model, which was an evolution of the SCSB model. The only difference from the ACSI model was that the NCSB model included the variable "corporate image" and its relationship to customer satisfaction and customer loyalty (Johnson et al., 2001). However, at the core of the model we find the relationship between quality, cumulative customer satisfaction, and customer loyalty, which has been measured since the very beginning of NCSB's existence, thus providing great insights for both the companies and researchers on how the level of customer satisfaction has developed over the years. Due to the shortcomings of the transaction-specific approach, the first version of the NCSB model was later expanded to include two relational dimensions, that is, calculative and affective commitment (ibid.). While the calculative commitment reflects customers' economic and rational reasons to continue the relationship, affective commitment reflects customers' warmer and more emotional motivations. Including commitment in the model led to a significant improvement in explaining customer loyalty. While the previous transaction-specific model explained only 20–30 percent of customer loyalty, the new model explained between 50 and 70 percent, depending on the type of industry (Johnson et al., 2001). In addition to including the relationship

dimensions, value was replaced by a purely price construct to avoid methodological problems between the earlier modeled quality and value variables. Recently, the NCSB model has been updated and extended by the addition of two variables (i.e., a company's digital solutions, and its sustainability and corporate social responsibility (CSR), reflecting priorities across industries). The current NCSB model can be seen in Figure 15.1.

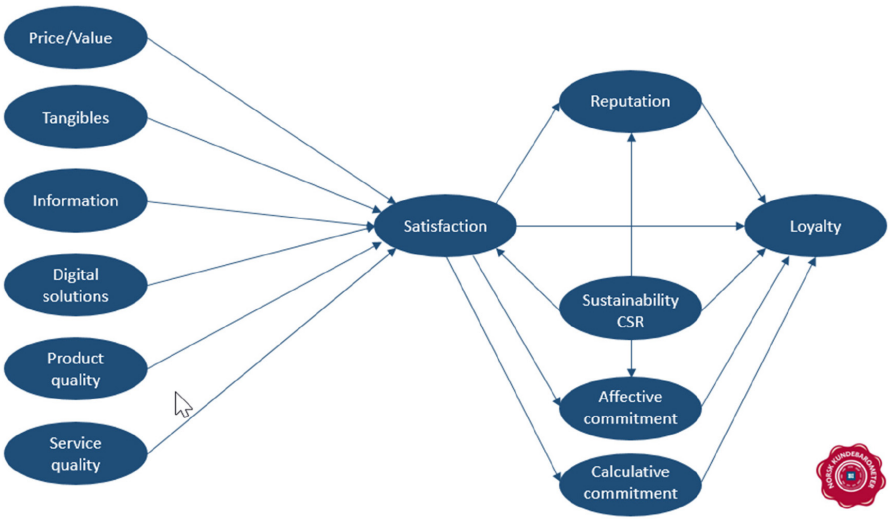

FIGURE 15.1. The Norwegian Customer Satisfaction Barometer Model 2018

The current NCSB model includes six dimensions that customers evaluate in a service encounter (i.e., the price/value, tangibles, information, digital solutions, product quality, and service quality) in determining their satisfaction. Satisfaction in turn is related to both types of commitment as well as loyalty. Furthermore, satisfaction helps building the company's reputation. The company's sustainability and CSR efforts play a central role in customers' evaluations of the company and affect satisfaction, affective commitment, reputation, and customer loyalty.

The evolution of the customer satisfaction barometer models does to a great extent reflect the evolution of three decades of service marketing research. First, applying the total quality management perspective to service marketing led to a focus on understanding quality in the 1980s. The focus shifted, however, to customer satisfaction in the 1990s as a response to the legitimate question, *Why is quality so important?* asked by both managers and academics. The answer of course being, *We want satisfied customers*. From around 1995 to 1999 the focus shifted to customer loyalty as a response to yet another legitimate question, that

is, *Why do we need to have satisfied customers?* The answer was, *Because we want loyal customers that come back, spread positive word of mouth and recommend the company to family and friends* (Zeithaml, Berry, & Parasuraman, 1996). At that time, it was established that it is less expensive to maintain existing customers than to constantly hunt for new ones. However, applying both strategies would be necessary, because the customer base is a leaky bucket.

While the customer satisfaction barometer models grasp the quintessence of the service encounter, they do not really include what is going on inside the company and how the inside affects the outside in terms of customer loyalty and profitability. The service-profit chain established the relationship between employee satisfaction on the inside and growth and profitability on the outside (Heskett, Jones, Loveman, Sasser, & Schlesinger,1994). More specifically, the logic of this model is that the internal service quality (i.e., the work environment) influences employees' satisfaction, productivity, and loyalty. Satisfied employees will be more productive and loyal, they will create greater value for customers, thus leading to customer satisfaction. Customer satisfaction drives loyalty, which ultimately affects both the company's growth and its profitability. This model has had a tremendous influence both in academia and for practitioners. For example, educational programs have been developed in line with the logic, while leaders across industries have changed their practices accordingly.

15.4 THE RELATIONSHIP PERSPECTIVE

The next phase in the evolution of the field is to gain a better understanding of what constitutes a relationship between organizations and their customers at an individual level. Customer satisfaction is without a question very important. Not all relationships, however, are created equal and organizations must differentiate how a customer is treated according to how value is created. Furthermore, an organization must link value creation within individual relationships to the overall value creation of the organization (Johnson & Selnes, 2004). It can be argued that an organization cannot survive in the long term having only customers that are extremely satisfied and loyal, as this base will slowly shrink and become smaller. It has been said that an organization's customer base is a leaky bucket and new customers constantly need to be added to cover the fixed costs. These ideas are closely related to the notion that all markets are created; no market is predefined. What companies must do is to nourish their market, or ecosystem, to make it grow and evolve continuously.

Also, although customer relationships with organizations are primarily economic relationships, they do have social meaning. Johnson and Selnes (2004) argue that customers start by being acquaintances and then are moved to being friends and then on to being partners. In this process, customers' perceptions of commitment and trust are altered. Another premise is that all customer relationships are important for an organization because they need them in the short term to cover the fixed cost and in the long term they need loyal customers. The underlying logic also dictates that organizations can differentiate their investments in the customer base according to the value of a customer. One example of a study that is built on this approach can be found in Tarasi, Bolton, Gustafsson, and Walker (2013). They show that customers having higher cash-flow levels (e.g., younger, purchasing more products in more categories) have higher variability in their cash flow. When variability is higher purchase patterns are more difficult to predict. Younger customers are also easier to acquire and lose, and therefore more unpredictable, while older customers could be more reliable and predictable, and therefore easier to serve and plan for. These characteristics help managers decide how to allocate sales and service efforts to segments.

As readers of this chapter will see later on, this evolutionary track does continue and is likely to arrive at with what is called the Internet of Things (we will return to this). However, before we do so, we need to cover another phase of the development: the customer experience perspective. This shift in focus was in part driven by Vargo and Lusch (2004), who suggested a revised logic focusing where services (rather than goods) are fundamental to economic exchange, and where the cocreation of value is the objective, thus emphasizing the role of the customer in any relationship. The experience is considered essential to value determination and the ecosystem is considered an active party in service provision (Lusch, Vargo, & O'Brien, 2007).

15.5 THE EXPERIENCE PERSPECTIVE

As physical store environments have started to have increased competition from online shopping, there has been an increased interest in the service encounter but from a slightly different perspective. Physical stores cannot compete head on with online stores on *price* therefore they must compete on something else—which happens to be the customer experience. The customer experience is defined as the period during which all service encounters relevant to a core service offering may occur (Voorhees et al., 2017). The service experience can be seen as a process while the encounter is a specific occurrence. The service experience is holistic in

nature and includes cognitive, affective, emotional, social, and physical responses to interactions with a service provider (Berry, Carbone, & Haeckel, 2002; Lemke, Clark, & Wilson, 2011).

There are at least three aspects of importance that we would like to point out. The *first* aspect is the process perspective; all service encounters imply a series of encounters that do not have to be with only one service provider (Lemon & Verhoef, 2016). What happens prior to and after the core encounter will affect the service experience. For instance, if a customer is given a coupon to use or even a recommendation from a friend, this action will influence downstream customer behavior and purchase decisions. Furthermore, customer processes may involve the use of different channels or touchpoints (i.e., Internet, catalogue, customer service, and a physical environment). It is very likely that customers will use different channels or touchpoints to achieve different goals. Because customers behave differently in different channels, companies would like to influence customers' choice of channels. For instance, if customers buy products online, they are less likely to make spontaneous purchases. To capture this process perspective, companies need to understand the *customer journey*. We will explore this question in more detail below.

The *second* aspect we want to highlight based on the definition of "service experience" is that it is a multi-dimensional, or holistic, perspective. As pointed out by Bitner (1992) in her conceptualization of servicescapes, there are a multitude of attributes that will influence customer perceptions of an experience both in an online environment and in a physical environment. The layout, colors, scents, symbols, or sounds as well as employee behavior and the possibility for customer interaction (with all aspects of a company) are important components that all influence the perception of service quality. For instance, the color of a room will influence how a wine tastes (Spence, Velasco, & Knoeferle, 2014); scent influences the perception of quality (Baker, Grewal, & Parasuraman, 1994). As seen from the definition, customer experience cannot be measured as an outcome of one dimension (e.g., satisfaction). The definition implies that aspects such as emotions or even physical responses may be important. Technological developments such as facial recognition make it possible to capture these types of outcome variables.

The *third* aspect which is difficult to capture in order to evaluate customer experience is the influence of a context. No company acts in a vacuum; customer experiences are formed by previous encounters and existing offerings from competitors. The experience is thus formed as a relative measure according to previous experiences of what competition is doing. Furthermore, what customers value var-

ies according to both demographics and culture. If a company is active in different markets, it has to take this variation into consideration. Companies such as McDonald's are well aware of this variation as shown by the fact that they adapt their offerings for different contexts.

15.5.1 THE CUSTOMER JOURNEY PERSPECTIVE

An experience can be thought of as the result of a process or as a customer journey that builds on multiple encounters at different touchpoints. Touchpoints can be seen as various ways a service provider has set up to interact with their customers. One of the ways to understand customer experiences is to follow in the footsteps of a customer across the touchpoints. On the Internet this is regularly done when customers leave traces on webpages they visit and by what they click on before making a purchase. This means companies can predict how customers will act in different situations. In a physical environment, predicting is slightly more difficult. Traditionally, this has been done by following customers around in stores and making notes on what they do. It looks as though new technology will make it possible to follow customers online just as easy as following them in a physical environment. If this happens, we will see predictions in real time in physical environments too. Personalized advertising can be used to nudge customers in making decisions just like in the movie the Minority Report where advertising is asking the main character John Anderton whether what he purchased the last time fits.

What is possible today is that a physical store can follow a customer throughout the store by communicating with the customer's cell phone without the customer even knowing (Henry, 2013). How this happens is that a cell phone is constantly looking for free Wi-Fi and a receiver can pick up that signal and follow the customer (cell phone) around a store. In relation to this, stores especially in Asia have started to implement facial recognition to understand who (e.g., age and gender) is doing the shopping in a store. Digital advertising may also include eye trackers to detect interest in the transmitted advertising; if a customer is not interested, a new advertisement will be displayed. In addition, technology is also being developed that in essence involves transmitters (e.g., Ibeacons) that are designed to send messages to customers who are using smartphones.

From a research perspective, these are very interesting technical solutions that will help us understand what a customer is doing in different settings. Customer engagement and connecting with a service provider emotionally are ways that service providers are trying to insulate customers from doing business on a transactional basis and/or just picking the best option according to price. With these

devices, we may understand what triggers arousal in different situations using, for instance, ways to measure electrodermal activity (EDA). We will also be able to understand what really happens with customer attention and behavior when scent or music is being used in physical stores.

This development is of course also slightly scary to think about but the technological solutions to accomplish this already exist and are already being tested or being used and we will have to get used to them in the future. What organizations are doing to counterbalance these opinions is to somehow pay customers to give away information.

15.6 INTERNET OF THINGS (IOT)

Technology is a game changer in the service context and it is changing how people behave (Larivière et al., 2017). For instance, new technologies such as smart grids, home management systems, electric cars, solar voltaic panels, and home batteries are changing the way customers perceive and manage energy consumption. Prerequisites of this technology are that Internet connectivity can be collected from all of these devices. This is the notion of the Internet of Things (IoT); everything is connected and information can be harvested. Systems may react autonomously. The IoT is a game changer in itself with the potential to affect consumers, businesses, and societies in unforeseen ways. The IoT is a network of entities that are connected through any form of sensor, enabling these entities, which we term as "Internet-connected constituents," to be located, identified, and even operated upon (Ng & Wakenshaw, 2017). The IoT represents a new context of service, characterized by a many-to-many, interconnected world, where people and devices are empowered by a constant flow of information and by the results of data analytics. The IoT is likely to disrupt many different markets such as healthcare, transportation, and retail.

As has recently been shown in the Cambridge Analytica scandal, data from the Internet, and in the future from the IoT, can potentially be linked to customers and used to predict when and where customers will want to have something, or will want to be connected to customer relationships. Here lies a potential danger but also a tempting future for the field—and this use of data is already being implemented. It was revealed recently that Amazon.com has obtained a patent for what it calls "anticipatory shipping"—a system of delivering products to customers before they place an order.

15.7 TRANSFORMATIVE SERVICE RESEARCH

Since 2010, researchers have started to recognize that because we are surrounded by all kinds of services every day, services actually have a much greater effect on our lives and well-being than traditional service dependent measures such as service quality, customer satisfaction, and loyalty reflect (Anderson & Ostrom, 2015). We go to school, to the doctor, we use financial services, governmental services, health care services, and more. The quality of these services affects our lives way beyond satisfaction and loyalty, they actually affect the quality of our lives and general well-being. This field of research is referred to as "transformative service research" and can be defined as research that focuses on "creating uplifting changes aimed at improving the lives of individuals, both consumers and employees, families, communities, society and the ecosystem more broadly" (Anderson & Ostrom, 2015, p. 243). In this area, both detractors and enhancers of well-being are studied in a variety of contexts, from financial services to health care. So far, topics such as cocreation, employee well-being, vulnerable consumers, social support, access, service literacy, service design, and service systems have been studied, while new themes are added as we speak (Anderson & Ostrom, 2015, p. 243).

15.8 FUTURE RESEARCH AVENUES

The service sector continues to grow in size and importance for any developed country. With the growth comes an endless interest in knowledge about the customer perspective of things, which is at the heart of service research. We want to generate information that helps customers receive a better experience or improve their way of life. This thirst for customer information seems to be endless at the moment. Companies such as Facebook, Skype, and Google have knowledge about pretty much every aspect of what happens in their customers' lives. The reason is that companies want to draw inferences from customer information in order to predict future behavior. Amazon, for instance, is predicting where a customer will make a purchase decision and send products ahead of time to ensure a quick delivery and a good customer experience. The IoT also makes it possible to localize and communicate with any product, and it will generate even more user information. Organizations will communicate with customers and employees in any physical environment, not just online. The communication is likely to occur with transmitters to communicate with various devices for the purpose of making customers aware of offerings in the vicinity. We have already started to see interactive advertising that reads faces and tries to figure out what generates customer interest. The

first self-checkout story has already been opened, meaning that every interaction a customer has with any product will be supervised. All of these developments are a bit scary to think about: so much information will be easily available and we must consider the implications. There is not much we can do to prevent such developments; it would be like trying to prevent the rain. What we can do is develop awareness and build on the benefits of this evolution.

REFERENCES

Anderson, E. W., Fornell, C., & Lehmann, D. L. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *The Journal of Marketing*, Vol. *58*(3), 53–66.

Anderson, L., & Ostrom, A. L. (2015). Transformative Service Research: Advancing Our Knowledge About Service and Well-Being, *Journal of Service Research*, Vol. *18*(3), 243–249.

Baker, J., Grewal, D., & Parasuraman, A. (1994). The influence of store environment on quality inferences and store image. *Journal of the Academy of Marketing Science*, *22*(4), 328–339.

Benoit, S., Baker, T. L., Bolton, R. N., Gruber, T., & Kandampully, J. (2017). A triadic framework for collaborative consumption (CC): Motives, activities and resources & capabilities of actors. *Journal of Business Research*, Vol. *79*(October), 219–227.

Berry, L. L., Carbone, L. P., & Haeckel, S. H. (2002). Managing the total customer experience, *MIT Sloan Management Review,* Vol. *43*(2), 85–89.

Bitner, M. J. (1992). Servicescapes: The impact of physical surroundings on customers and employees. *The Journal of Marketing*, 57–71.

Fornell, C. (1992). A national customer satisfaction barometer: The Swedish experience, *Journal of Marketing,* Vol. *56*(1), 6–21.

Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., & Bryant, B. E. (1996). The American customer satisfaction index: nature, purpose, and findings. *Journal of Marketing*, Vol. *60*(4), 7–18.

Fornell, C., Mithas, S., Morgeson III, F. V., & Krishnan, M. S. (2006). Customer satisfaction and stock prices: High returns, low risk. *Journal of Marketing*, *70*(1), 3–14.

Grönroos, C. (1998). Marketing Services: The Case of a Missing Product. *Journal of Business and Industrial Marketing,* Vol. *13*(4/5), 322–338.

Henry, A. (2013). How Retail Stores Track You Using Your Smartphone (and How to Stop It), Lifehack available at https://lifehacker.com/how-retail-stores-track-you-using-your-smartphone-and-827512308, accessed January 29, 2018.

Heskett, J. L., Jones, T. O., Loveman, G. W., Sasser, W. E. Jr., & Schlesinger, L. A. (1994). Putting the Service-Profit Chain to Work. *Harvard Business Review,* Vol. *72*(2), 164–174.

Johnson, M. D., & Fornell, C. (1991). A framework for comparing customer satisfaction across individuals and product categories. *Journal of Economic Psychology,* Vol. *12*(2), 267–286.

Johnson, M. D., & Gustafsson, A. (2000). *Improving Customer Satisfaction, Loyalty and Profit: An Integrated Measurement and Management System*, San Francisco, CA: Jossey-Bass.

Johnson, M. D., Anderson, E. W., & Fornell, C. (1995). Rational and adaptive performance expectations in a customer satisfaction framework. *Journal of Consumer Research,* Vol. *21*(4), 665–707.

Johnson, M. D., & Selnes, F. (2004). Customer portfolio management: toward a dynamic theory of exchange relationships. *Journal of Marketing*, Vol. *68*(2), 1–17.

Johnson, M. D., Gustafsson, A., Andreassen, T. W., Lervik, L., & Cha, J. (2001). The evolution and future of national customer satisfaction index models. *Journal of Economic Psychology*, Vol. *22*(2), 217–245.

Kumar, V., & Pansari, A. (2016). National Culture, Economy, and Customer Lifetime Value: Assessing the Relative Impact of the Drivers of Customer Lifetime Value for a Global Retailer. *Journal of International Marketing,* Vol. *24*(1), 1–21.

Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Siriannie, N. J., Voss, C., Wünderlich, N. V., & De Keyserh, A. (2017). Service Encounter 2.0: An Investigation into the Roles of Technology, Employees and Customers, *Journal of Business Research*, Vol. *79*(October), 238–46.

Lemke, F., Clark, M., & Wilson, H. (2011). Customer experience quality: An exploration in business and consumer contexts using repertory grid technique. *Journal of the Academy of Marketing Science,* Vol. *39*(6), 846–869.

Lemon, K. L., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, Vol. *80*(1), 69–96.

Lusch, R. F., Vargo, S. L., & O'Brien, M. (2007). Competing through service: Insights from service-dominant logic. *Journal of Retailing*, Vol. *83*(1), 5–18.

Ng, I. C., & Wakenshaw, S. Y. (2017). The Internet-of-Things: Review and research directions. *International Journal of Research in Marketing*, *34*(1), 3–21.

Ostrom, A. L., Bitner, M. J., Brown, S. W., Burkhard, K. A., Goul, M., Smith-Daniels, V., & Rabinovich, E. (2010). Moving forward and making a difference: research priorities for the science of service. *Journal of Service Research*, *13*(1), 4–36.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing,* Vol. *49*(4), 41–50.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). Servqual: A multiple-item scale for measuring consumer perception. *Journal of Retailing*, *64*(1), 12–40.

Rego, L. L., Morgan, N. A., & Fornell, C. (2013). Reexamining the market share–customer satisfaction relationship. *Journal of Marketing*, Vol. *77*(5), 1–20.

Rust, R. T., Moorman, C., & Dickson, P. R. (2002). Getting Return on Quality: Revenue Expansion, Cost Reduction or Both? *Journal of Marketing*, Vol. *66*(4), 7–24.

Shostack, L. G. (1977). Breaking Free from Product Marketing. *Journal of Marketing*, Vol. *41*(2), 73–80.

Shostack, L. G. (1985). Planning the Service Encounter. In *The Service Encounter*. J. A. Czepiel, M. R. Solomon, & C. F. Surprenant (Eds.), New York: Lexington Books, pp. 243–254.

Spence, C., Velasco, C., & Knoeferle, K. (2014). A large sample study on the influence of the multisensory environment on the wine drinking experience. *Flavour*, 3:8.

Tarasi, C. O., Bolton, R. N., Gustafsson, A., & Walker, B. A. (2013). Relationship characteristics and cash flow variability: implications for satisfaction, loyalty, and customer portfolio management. *Journal of Service Research,* Vol. *16*(2), 121–137.

Vargo, S. L., & Lusch, R. F. (2004). Evolving to a New Dominant Logic for Marketing. *Journal of Marketing*, *68*(1), 1–17.

Voorhees, C. M., Walkowiak, T., Fombelle, P. W., Gregoire, Y., Bone, S., Gustafsson, A., & Sousa, R. (2017). Service encounters, experiences and the customer journey: Defining the field and a call to expand our lens. *Journal of Business Research,* Vol. *79*(October), 269–280.

Zeithaml, V. A. (1981). How Consumer Evaluation Processes Differ between Goods and Services. In *Marketing of Services.* J. Donnelly & W. George (Eds.), Chicago: American Marketing, pp.186–190.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The Behavioral Consequences of Service Quality. *Journal of Marketing,* Vol. *60*(2), 31–46.

Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1985). Problems and Strategies in Services Marketing. *Journal of Marketing,* Vol. *49*(2), 33–46.

Biographies

CHAPTER 1

Amir Sasson: Amir Sasson is a provost and professor at BI Norwegian Business School. Amir Sasson has a PhD from BI, Norwegian Business School. He has been a lecturer at University College Dublin, Ireland, and a visiting scholar at Stanford University, USA, and The Technion, IL. He has published in top journals in the fields of network theory, inter-organizational relations, entrepreneurship, business models and industrial competitiveness. He has recently co-headed a large national research project on the competitiveness of Norwegian industries, Knowledge-Based Norway.

CHAPTER 2

Espen Andersen: Espen Andersen is Associate Professor at the Department of Strategy and Entrepreneurship at BI Norwegian Business School, where he co-leads the Center for Digitalization, and Associate Professor II at the Department of Computer Science at the University of Oslo. His research interests include digitalization, technology strategy, technology evolution, cognitive technologies and strategy in technology-intensive industries, and case teaching. Andersen is a regular commentator on technical and strategic questions in the media, and has been a consultant, teacher and researcher on four continents. He holds the degree of Doctor of Business Administration from Harvard Business School. For more information, see Andersen's personal website, espen.com.

John Chandler Johnson: John Chandler Johnson holds a PhD in Sociology from Stanford University, and is an Associate Professor at BI Norwegian Business School. His research formalizes microsociological models of endogenous network evolution to simulate disequilibrium knowledge dynamics emergent in processes of social structural reproduction.

Vegard Kolbjørnsrud: Vegard Kolbjørnsrud is an Assistant Professor in Strategy at BI Norwegian Business School and a Senior Research Fellow (Adj.) at Accenture. His research interests include governance and organization design, communities and crowdsourcing, and the impact of technology on strategy and management. Vegard has recently published in journals such as *Harvard Business*

Review and *Strategic Organization* and his research has been covered in national and international news media. Prior to joining BI Kolbjørnsrud was a Senior Manager in Strategy at Accenture and a Senior Research Fellow at Accenture Institute for High Performance. He received his PhD in Strategy from BI Norwegian Business School in Oslo, Norway.

Ragnvald Sannes: Ragnvald Sannes is Senior Lecturer at the Department of Strategy and Entrepreneurship at BI Norwegian Business School, where he co-leads the Center for Digitalization. His research interests include digitalization and digital transformation in combination with new business models and practices, strategic change, and organizational issues. Sannes is an experienced advisor on technology and innovation and has coached more than 200 innovation projects in Norwegian organizations. Ragnvald is an active member of the Norwegian information systems community. He leads the National Academic Council for Information Science and the Norwegian Computer Society's special interest group on Strategy and Management. Ragnvald is a regular commentator on technology and innovation, as an educator, researcher, advisor and speaker. He holds a Licentiate degree in Business Administration from Stockholm School of Economics.

CHAPTER 3

Sut I Wong: Sut I Wong, PhD, is currently a Professor in Communication and Leadership at BI Norwegian Business School and an adjunct at University of Ljubljana. She is also a director at the Nordic Centre for Internet and Society. She received her DBA in management from Catholic University of Portugal in 2008 and PhD in organizational psychology from BI Norwegian Business School in 2013. Her research largely focuses on relationship-based leadership, virtual team leadership, and human resource practices for individual innovative and proactive behaviors, such as job crafting. Her more recent research looks into management practices for digital change with projects funded by the Norwegian Research Council. Her research has been published in top-tier international journals, such as *Journal of Management*, the *Leadership Quarterly*, *Human Resource Management*, *Human Resource Management Journal*, and *Journal of Occupational and Organizational Psychology*.

Christian Fieseler: Christian Fieseler is Professor of Communication Management at BI Norwegian Business School and a director at the Nordic Centre for Internet and Society. He received his PhD in Management and Economics from the University of St. Gallen, Switzerland, in 2008. Since then, his research has

focused on the question of how individuals and organizations adapt to the shifts brought about by new, social media, and how to design participative and inclusive spaces in this new media regime. In this field, he has worked extensively in recent years on technology and new working modes in projects with the European Union and the Norwegian Research Council.

CHAPTER 4

Peggy Simcic Brønn: Peggy Simcic Brønn is Professor of Communication and Management at the Norwegian Business School's Department of Communication and Culture and director of BI's Center for Corporate Communication. Her most recent co-authored publications on reputation include "Systems Thinking: A Method for Reducing Reputation Risk" (in *International Studies of Management and Organization*), and "A good reputation: Protection against shareholder activism" (in *Corporate Reputation Review*).

Alexander Buhmann: Alexander Buhmann is an Assistant Professor at the Department of Communication and Culture at BI Norwegian Business School and serves as the Co-Director of the BI Centre for Corporate Communication. He is also a Research Fellow at the Annenberg School for Communication and Journalism and University of Southern California Center on Public Diplomacy. Alexander holds an MA in Media Studies from the University of Siegen, and a PhD in communication from the University of Fribourg. His current research interests include communication management, digitalization, and public diplomacy.

CHAPTER 5

Lars Erling Olsen: Professor Olsen teaches marketing at BI Norwegian Business School. He researches brand strategy, marketing communication and persuasion theory. His research has been published in journals such as *Journal of Business Research*, *Journal of Advertising*, *Marketing Letters*, *Psychology & Marketing*, and *European Journal of Marketing*. He is co-author of the first Norwegian academic textbook on brand management, and recently a book on marketing communication.

CHAPTER 6

Øystein D. Fjeldstad: Professor Fjeldstad teaches strategy at BI Norwegian Business School. He researches value creation, business models, and collaborative organization design. His research has been published in journals such as *Strategic Management Journal*, *Strategic Organization*, *Long Range Planning*, *Journal of Management Studies*, and *Social Networks*. Fjeldstad's work draws on studies of telecommunications, financial services, technology, and shipping.

Knut Haanæs: Knut Haanæs is the Dean of the Global Leadership Institute at the World Economic Forum, and Professor of Strategy at IMD. He is formerly a Senior Partner in The Boston Consulting Group and has also been an Associate Professor of Strategy at BI Norwegian Business School. He has published in journals such as *Harvard Business Review* and *MIT Sloan Management Review*, and in 2015 co-authored *Your Strategy Needs a Strategy*, which has been translated into seven languages.

CHAPTER 7

Janis Berzins: Janis Berzins researches corporate finance and governance issues pertaining to non-listed family firms and works on projects spanning profitability, owner-firm liquidity shock propagation, agency issues, and payout policy. He also researches strategy, performance attribution, and information processing in mutual fund and institutional money management complexes.

Øyvind Bøhren: Professor Bøhren works with corporate finance in general and corporate governance in particular (http://home.bi.no/oyvind.bohren). His current research projects analyze the relationship between ownership structure, taxes, and dividends, the effect of shareholder illiquidity on firm behavior, the merits of family management in family firms, and the performance premium of family firms. Bøhren is the founding director of the Centre for Corporate Governance Research (www.bi.edu/ccgr), which was established in 2005.

Bogdan Stacescu: Bogdan Stacescu works in corporate governance, corporate finance, and banking. His current research projects examine the performance of family firms, the links between personal and corporate financial constraints, the size of credit bureaus, and the use of collateral as a selection device by banks.

CHAPTER 8

Tore Bråthen: Tore Bråthen is Professor of Business Law at BI Norwegian Business School and works especially with legal questions related to state ownership, company law, contact law and real estate agents law. He is the editor of *Nordisk Tidsskrift for Selskabsret* (NTS, Nordic Journal of Company Law), *Norsk Lovkommentar* (Norwegian Commentary on the Law) and he is a member of the editorial board of *Tidsskrift for Eiendomsrett* (Journal of Real Estate Law) and *Corporate Ownership and Control*. He has acted as chairman for several legislative committees appointed by the government. He has written numerous scholarly articles and books.

CHAPTER 9

Sverre August Christensen: Sverre August Christensen is Associate Professor in Business History, with a Dr. Oecon from the Norwegian Business school, awarded in 2006. He has several publications on Norwegian and international business history. He is particularly interested in different forms of ownership, state, national and international ownership, and its impact on industrial development and corporate governance.

CHAPTER 10

John Christian Langli: John Christian Langli holds a PhD from the Norwegian School of Economics (NHH) and is currently Professor of Business Economics at the Department of Accounting, Auditing and Business Analytics at BI Norwegian Business School. His research interests are empirical research in financial accounting, auditing and corporate governance.

Marleen Willekens: Marleen Willekens holds a PhD from the University of Warwick. She is currently Full Professor of Accounting and Auditing (KU Leuven, Belgium) and Adjunct Professor of Auditing at BI Norwegian Business School. She has previously been Professor of Auditing at Tilburg University (the Netherlands) and member of the executive body and the general body of the KU Leuven Research Council. Her research interest is empirical and quantitative research in the field of accounting and auditing.

CHAPTER 11

Miha Škerlavaj: Miha Škerlavaj is a Professor at BI Norwegian Business School, Department for Leadership and Organizational Behavior. His research interests include creativity, innovation, change, knowledge hiding and prosocial behaviors at work. His work has been published in top academic journals, including the *Academy of Management Journal*, the *Leadership Quarterly* and the *Journal of Organizational Behavior*.

CHAPTER 12

Bård Kuvaas: Bård Kuvaas is Professor of Organizational Psychology at the Department of Leadership and Organizational Behavior at BI Norwegian Business School. He received the BI Norwegian Business School Research Prize for 2016, a prize that is awarded every three years.

CHAPTER 13

Geir H. M. Bjertnæs: Geir H. M. Bjertnæs has a Cand. Oecon. from the University of Oslo. He is currently working as a researcher at Statistics Norway. He was project manager of the leading CGE model in Norway from 2010 to 2017. He has extensive experience working with theoretical research questions related to environmental economics and public policy. He has published in several journals, including *Energy Economics*, *Energy Policy* and *Public Finance Review*.

Tom-Reiel Heggedal: Tom-Reiel Heggedal has a PhD in Economics from the University of Oslo and is currently Associate Professor of Economics at BI Norwegian Business School. He has extensive experience working with theoretical research questions related to applied game theory, environmental economics, labor markets, and public policy. He has published in top journals in these fields, including the *Journal of Economic Theory*, the *Journal of Economic Behavior and Organization*, and the *Journal of Public Economics*.

Karl Jacobsen: Karl Jacobsen has an MSc in Economics from the University of Bergen and an MSc in Computational Science and Engineering from the University of Oslo. He has published in journals such as *Energy Economics* and *The Scandinavian Journal of Economics*.

CHAPTER 14

Leif Helland: Leif Helland holds a Dr. Polit. degree from the University of Oslo and is Professor of Economics at BI Norwegian Business School. He is also director of BI's Centre for Experimental Studies and Research (CESAR). He works in applied game theory, experimental economics, and political economy. Helland has published in top journals within economics and political science, such as the *Journal of Economic Theory*, *Journal of Theoretical Politics, Journal of Economic Behavior and Organization*, *Electoral Studies*, and *Journal of Conflict Resolution*.

Lars Chr. Monkerud: Lars holds a Dr. Econ. degree from BI Norwegian Business School. He currently works as a researcher at OsloMet. His research falls mainly within health economics, political economy, and local government studies. Monkerud has published his work in top journals such as *Journal of Theoretical Politics*, *European Journal of Political Research*, *Health Economics*, and *Human Reproduction*.

Gjermund Løyning: Gjermund holds an MSc in Political Economy from BI Norwegian Business School. He is currently a Deputy Director of Government Relations at the Confederation of Norwegian Enterprise (NHO). Gjermund has previously worked with government relations in Energy Norway, and was a political advisor to the Minister for Church and Culture in the second Bondevik government.

CHAPTER 15

Anders Gustafsson: Anders Gustafsson is a Professor of Business Administration in the Service Research Center at Karlstad Business School in Sweden. He holds a part-time position at Norwegian Business School. He is the current editor-in-chief for *Journal of Business Research* and an area editor for *Journal of Service Research*. Gustafsson has published articles in journals such as *Journal of Marketing*, *Journal of Marketing Research*, *Journal of Service Research*, *Journal of Business Research*, and *Industrial Marketing Management*. Gustafsson is also the author of 12 books.

Line Lervik-Olsen: Line Lervik-Olsen is a Professor of Marketing at BI Norwegian Business School and holds a part-time position at the Norwegian School of Economics and the Center for Service Innovation. Her work has been published in journals such as the *Journal of Service Research*, the *Journal of Economic*

Psychology, *Journal of Service Management*, *Managing Service Quality*, *Journal of Services Marketing*, *Journal of Service Theory and Practice*, *PLOS One*, *Journal of Business Research*, as well as in various books.