THE CAPCO INSTITUTE
JOURNAL
OF FINANCIAL TRANSFORMATION

DESIGN THINKING

#48 NOVEMBER 2018

THE CAPCO INSTITUTE

JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

CONTENTS

DESIGN

TRANSFORMATION

SUPERVISION

DEAR READER,

Design thinking, a collaborative, human-focused approach to problem-solving, is no longer just for the creative industries. It has become an important management trend across many industries and has been embraced by many organizations. Its results are hard to ignore. Indeed, design-driven companies regularly outperform the S&P 500 by over 200 percent.[1]

To date, the financial services industry has not led in adopting this approach. However, leaders are recognizing that important challenges, such as engaging with millennial customers, can be best addressed by using design thinking, through the methodology's exploratory approach, human focus, and bias towards action. This edition of the Journal examines the value of design thinking in financial services.

Design thinking introduces a fundamental cultural shift that places people at the heart of problem-solving, which is critical in a technology-driven environment. If the customer's real problems are not fully understood, technological solutions may fail to deliver the desired impact. In this context, design thinking offers a faster and more effective approach to innovation and strategic transformation.

The case studies and success stores in this edition showcase the true value of design thinking in the real world, and how this approach is an essential competitive tool for firms looking to outperform their peers in an increasingly innovation-driven and customer-centric future. At Mastercard, design thinking has become a part of almost all organizational initiatives, from product development, research and employee engagement to solving challenges with customers and partners. Meanwhile, at DBS Bank in Singapore, a data-informed design model has been firmly embedded into the bank's culture, enabling them to successfully move from being ranked last among peers for customer service in 2009, to being named the Best Bank in the World by Global Finance in 2018.

I hope that you enjoy the quality of the expertise and points of view on offer in this edition, and I wish you every success for the remainder of the year.

Lance Levy, **Capco CEO**

[1] http://fortune.com/2017/08/31/the-design-value-index-shows-what-design-thinking-is-worth/

DESIGN

DESIGN THINKING AS A PROCESS FOR PEOPLE-CENTERED INNOVATION IN THE FINANCIAL SECTOR[1]

RAMA GHEERAWO | The Helen Hamlyn Centre for Design, Royal College of Art
JEREMY MYERSON | The Helen Hamlyn Centre for Design, Royal College of Art

ABSTRACT

This paper provides an overview of the origins, principles, values, and benefits of design thinking as a creative framework for innovation in business. It looks in particular at the rise of design thinking in the ten years since the global financial crash and speculates on the suitability of its methods for current transformation challenges in the financial services sector. In setting out practical frameworks for adoption, the paper presents three people-centered projects from the Helen Hamlyn Centre for Design at the Royal College of Art, carried out with financial companies.

1. INTRODUCTION

There has been growing momentum over the past ten years for non-designers to enter the design space – for engineers, entrepreneurs, managers, and social scientists in particular to think and act like designers by adopting 'designerly' ways of thinking and emulating the methods designers use to innovate. This movement has been given a name, "design thinking," and it has become evident across a broad span of design-based activity, from delivery of public services to development of consumer electronics. The impact in many business circles is tangible.

Design thinking is currently less prevalent in financial services, although that is starting to change and its core principles are relevant to the transformation journey that the sector is undergoing. That is because design thinking has proven to be an effective method for harnessing the creative, innovative, and people-centered approaches enshrined in the design process and applying them to organizational challenges. In effect, it provides an alternative lens through which to view business problems and identify solutions to solve them.

A catalyst for adoption of design thinking was a Harvard Business Review paper published in 2008 by Tim Brown, chief executive officer of the global innovation consulting firm IDEO. The article described design thinking as a tool for business, explaining it as a discipline that uses the designer's sensibility and methods to match people's needs and desires with what is technologically feasible and what is viable as a business strategy. In a parallel development, several business and engineering schools, led by Stanford University in Palo Alto, where IDEO is headquartered and Brown and his colleagues teach, set up 'D-Schools' to explore the topic. Tim Brown's follow-up book, Design for change: how design thinking transforms

[1] Design thinking is taught as an Executive Masterclass to companies at the Royal College of Art by Rama Gheerawo, Director of the Helen Hamlyn Centre for Design, and Jeremy Myerson, Helen Hamlyn Professor of Design.

organizations and inspires innovation in 2009, further broadened the debate on what design thinking actually means and how its key principles work in practice.

In its simplest form, to think like a designer involves a number of basic things, such as: showing human empathy with people rather than being scientifically neutral; adopting a participatory mindset rather than an expert one, so you can ask the dumb questions and challenge accepted wisdom; and making use of design tools and skills such as visualization and prototyping to share ideas, and elicit and incorporate feedback as part of a co-design process. A key question for the financial service industry, as it struggles to reconnect with customers and staff amid a crisis of legitimacy and trust following the global financial crash, is how much impact design thinking might have in a field where its influence to date has been less evident than in other spheres.

> "A definitive move that strengthens the relationship between design thinking and business is the development of design thinking methods, tools, teaching curricula, and practice at business schools."

2. BACKGROUND AND DEFINITION

While design thinking has gained currency in the decade directly following the crash, its practices predates its modern nomenclature. Some scholars argue that its ideas and ideals resonate across the centuries and even reach back to the achievements of ancient civilizations. Human history is thought to be full of design thinkers, long before the term was defined and popularized. Leonardo Da Vinci, Isambard Kingdom Brunel, and Richard Buckminster Fuller all used or created processes that resonate with a design thinking approach that modern practitioners would recognize – from a desire to better the human condition using a cross-disciplinary approach, to taking advantage of technological advances to invent and innovate in a way that tackles complex challenges situations. In his 2009 book, Tim Brown himself used Brunel's passenger-focused strategy for the engineering and design of his Great Western Railway from London to Bristol in 1841, with trains "floating across the countryside," as a pioneering early example of design thinking in action. To achieve such a user-centric effect, Brunel built bridges, viaducts, cuttings, and tunnels to create a smooth experience.

As to the more modern origins of design thinking as a phrase or a concept, Mottee (2013) and Chae (2017) suggest Herbert Simon's 1969 book, The sciences of the artificial, as the initial point at which design was proposed as a way of thinking, highlighting the definition of design as "the transformation of existing conditions into preferred ones." The book proposes seven steps which still relate to current processes of design thinking as follows: define, research, ideate, prototype, choose, implement, and learn.

Szczepanska (2017) talks about the Design Science movement that originated in the U.S. in the 1960s, where Buckminster Fuller created multidisciplinary design teams to address complex systemic challenges. Peter Rowe's 1987 book, Design thinking, is also notable, though its focus was on the process of designing in architecture and urban planning rather than design thinking as it is known today. Kleinsmann et al. (2017) cite another early use of the term from Bruce Archer in his 1979 article, "Whatever became of design methodology?" stating that design thinking originated within the design research community. Archer helped found the Design Research Society in 1967 and established the Department of Design Research at London's Royal College of Art (RCA), where The Helen Hamlyn Centre for Design (the host institute for the authors of this papers) is now based.

In today's context, design thinking has become a living, evolving idea that is being adopted by a range of individuals and organizations. It is also being defined and redefined and means different things to different people [Johansson-Skoldberg et al. (2013)]. Importantly, at the center of design thinking is the idea that everyone can access and activate the creative value of design. Cross (2011) states that "everyone can – and does – design," noting that humans have had a long history of design thinking as evidenced by the artefacts and inventions of previous civilizations and the "continuing tradition of vernacular design and traditional craftwork."

Exact definitions of design thinking vary but some commonalities and convergences have emerged. Dorst (2015) talks about design thinking as a "real alternative to conventional problem-solving strategies," something which design thinking case studies attest to. The move

to address systemic challenges through design thinking [Luchs (2016)] or create organizational change by establishing a design thinking culture within an institution or company [Calabretta et al. (2008)] are ideas that have gained visibility in the last decade. The people-centered aspect of design thinking is also prevalent, with the terms human-centered and user-centered used unilaterally within the field. Fraser (2012) talks about "deep human understanding" based on work at the Rotman School of Management. Curedale (2015) notes that design has moved from being a marketing tool to answering human need.

3. PRINCIPLES OF DESIGN THINKING

A definitive move that strengthens the relationship between design thinking and business is the development of design thinking methods, tools, teaching curricula, and practice at business schools. The d.school at Stanford University, founded in 2005,[2] outlines eight "core abilities":

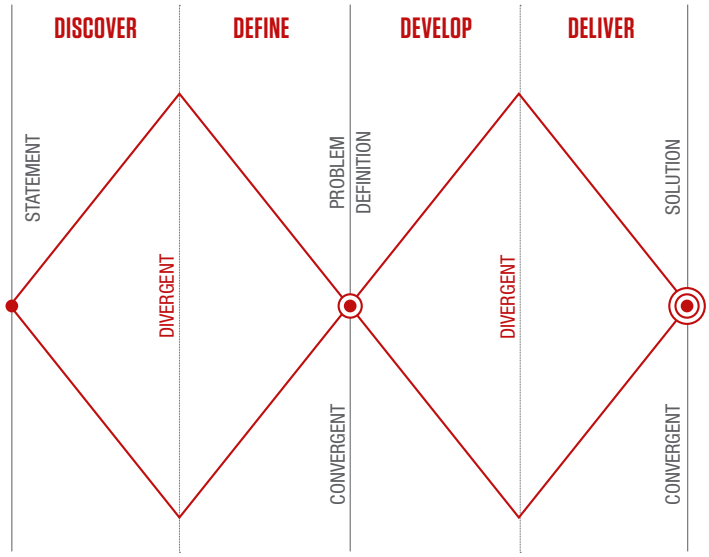
1. Navigate ambiguity
2. Learn from others (people and contexts)
3. Synthesize information
4. Experiment rapidly
5. Move between concrete and abstract
6. Build and craft intentionally
7. Communicate deliberately
8. Design your design work

An empirical interview study of five large organizations by Carlgren et al. (2016) led to five themes said to characterize design thinking: user focus, problem framing, visualization, experimentation, and diversity.

"User focus" is about "deep empathy building" as well as understanding and involving people in the generation and making of ideas. **"Problem framing"** looks at "widening, challenging, and reframing" a problem statement instead of simply trying to solve it, and avoiding the trap of narrowing down to a solution too quickly. **"Visualization"** aims to use "visual representations" to show ideas in either two or three dimensions to gain consensus and share ideas rather than simply being about prototyping and delivering a solution. **"Experimentation"** refers to iteratively developing and testing ideas in ways that are convergent and divergent, working on multiple solutions to maximize the creative value of process and outcome. **"Diversity"** was seen as a cross-cutting theme applicable to ensuring a range of opinions and perspectives as well as a diversity of team members.

[2] https://stanford.io/2yMGRcd

Figure 1: The Double Diamond model outlines the place of convergent and divergent thinking within the process

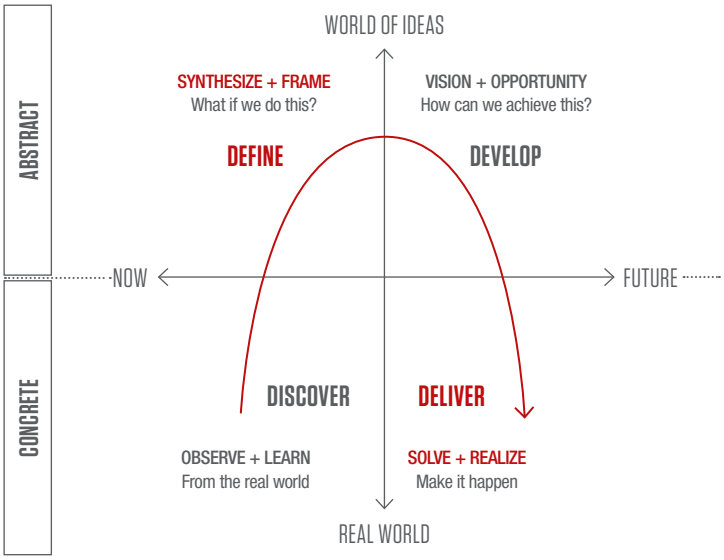


Source: U.K. Design Council

Within our own institution, The Helen Hamlyn Centre for Design, we have developed design thinking frameworks, tools, and practice over the past 10 years through close connection with IDEO, the U.K. Design Council, and others. These are our own principles for engagement with design thinking:

- **Look and learn**: a primary characteristic of design thinking is to be patient, to go into the field, to observe and record in a sketchbook or with a camera, and to do user research without preconceptions. So much business thinking is based on preconceived ideas, on existing market "knowledge," and an over-awareness of barriers to change. When asked to design a new product, service, or communication, designers look at things in a fresh and sometimes naïve way, asking the dumb questions and behaving like participants in a process, not experts. That way, they look and learn. Design ethnography need not be complicated – it can simply be about walking a mile in your customer's shoes.
- **Prototype early and often:** the prototype tends to mark the final stage before production in default business thinking: "this is what it is going to be like." But design thinkers treat the prototype differently. They build, test, and experiment in an iterative loop, revising from one prototype to the next to learn about what will work. Remember that James Dyson, a Royal College of Art graduate, experimented with more than five thousand prototypes before perfecting the dual-cyclone vacuum cleaner that built his business empire. So the key message is try things out first – and do not fret if they fail. As a business, you will discover a lot and succeed sooner by prototyping often.
- **Don't be afraid to cross-pollinate:** business managers are often specialists in a particular field – and their thinking is bounded by that field of expertise. But designers tend to take a more generalist approach that means lessons in one sector can be applied to another. One of the central tenets of design thinking is a willingness to cross-pollinate – to take ideas from one area and apply them in a totally different context. Can the pit stop tyre change process in Formula 1 racing be translated into the team dynamics in the accident & emergency department of a hospital? Can aerospace technology be inserted into an ergonomic office chair? Can the typography associated with prayer books help to sell shampoo? There can be surprising solutions to business problems if you are willing to be open-minded and cross-pollinate.
- **Think visually, not in words:** many professionals rationalize or justify design decisions by writing long reports with lots of words to wade through. Design thinkers use images. Their way of thinking is visual. Simple diagrams, photo-evidence, development sketches, etc., all help to communicate ideas and support effective and collaborative design decision-making. When the authors of this paper were asked to organize a design thinking seminar at 10 Downing Street for senior civil servants, one of the key learnings from the day was to put more images and less words into briefings for government ministers, as these had more impact in terms of argument and evidence, and saved time.
- **Know the limits of design thinking:** as a designer thinker, you can look and learn, you can prototype early and often as you test designs with users, you can cross-pollinate ideas from one sector to another, and you can think visually at every opportunity. But that does not make you a designer. The final principle is to know the limits of design thinking. Even when you have collected a wealth of customer evidence and formulated and tested your innovative new approach, you will still need the services of a professional designer to help make the project a reality in the marketplace. In our view, design thinking is a useful bridge between designers and those who commission and use design, a shared set of perspectives or values so that everyone is on the same page and pulling

Figure 2: design thinking manages a shift of focus between "now" and the "future," and the "world of ideas" and the "real world"



Source: The Helen Hamlyn Centre for Design, Royal College of Art

in the same direction when it comes to making the project a success.

4. FRAMEWORK FOR DESIGN THINKING

As well as establishing our principles, we have also established our own framework for using design thinking to address business problems. We took as our starting point the "Double Diamond" model of innovation developed by the U.K. Design Council (Figure 1). This sets out four key stages: discover, define, develop, and deliver. Each stage involves either convergent thinking or divergent thinking depending on whether you wish to create choices (blue sky thinking) or make choices (focusing ideas) within the process.

Typical activities in each phase are:

- **Discover phase (divergent):** exploration of brief and hypothesis, contextual research and definition of project participants or communities. Looking at the world in a novel way and gathering insights
- **Define phase (convergent):** designing and conducting design ethnography. People-centered briefs defined from the research insights. Curating and implementing the possibilities identified in the Discover phase. Writing of design brief that nails the real problem.
- **Develop phase (divergent):** development of a number of ideas through co-creation and design ideation processes. This process of trial and error helps to improve and refine ideas.
- **Deliver phase (convergent):** selection of ideas to take forward and delivering outputs in the form of prototypes, services ideas, or guidance. The resulting project is finalized, produced, and launched

However, our own framework adds a new dimension by negotiating the relationship between the real world (concrete) and the world of ideas (abstract), and by placing the define and develop stages in the world of ideas, not the real world, thus freeing up innovation potential. Figure 2 sets out the process.

The four quadrants that result from the framework described in Figure 2 lend themselves to the following design thinking activities, creating an "arc of design thinking" from "discover" (observe and learn), through "define" (synthesis and frame), and "develop" (vision and opportunity) to "deliver" (solve and realize).

- **Now-real world:** this is about observing and learning from what is currently happening. Activities are about understanding context and people's behaviors in existing situations and gathering insights that can be evaluated. This takes place in the present moment and builds up a real picture of any given situation.
- **Now-world of ideas:** synthesizing and reframing the insights that were gathered in the previous stage is the main activity here. Asking questions such as "what if we do this?" talks to the speculative characteristic of this. This is a transformational process that moves a project into more imaginative and unknown spaces.
- **Future-world of ideas:** here, the focus is on articulating a vision and defining opportunities. Design briefs are typically formulated and answered at this point and a number of creative avenues are delineated and explored. It is important to nourish every idea at this stage as often the most creative ideas come from unfettered ideation.
- **Future-real world:** the final stage is about solving, realizing, and delivering ideas back in the real world, making sure that they are relevant to the people who will most benefit from them, and that they are market-appropriate. Activities such as prototyping and evaluation often take place at this point.

When we map "The arc of design thinking" onto the Double Diamond model, our final model can be described visually as that presented in Figure 3.

5. DESIGN THINKING CASE STUDIES IN THE FINANCIAL SERVICES SECTOR

The Helen Hamlyn Centre for Design at the Royal College of Art has used this framework to work with partners to address innovation challenges in the financial services sector. Three are discussed here.

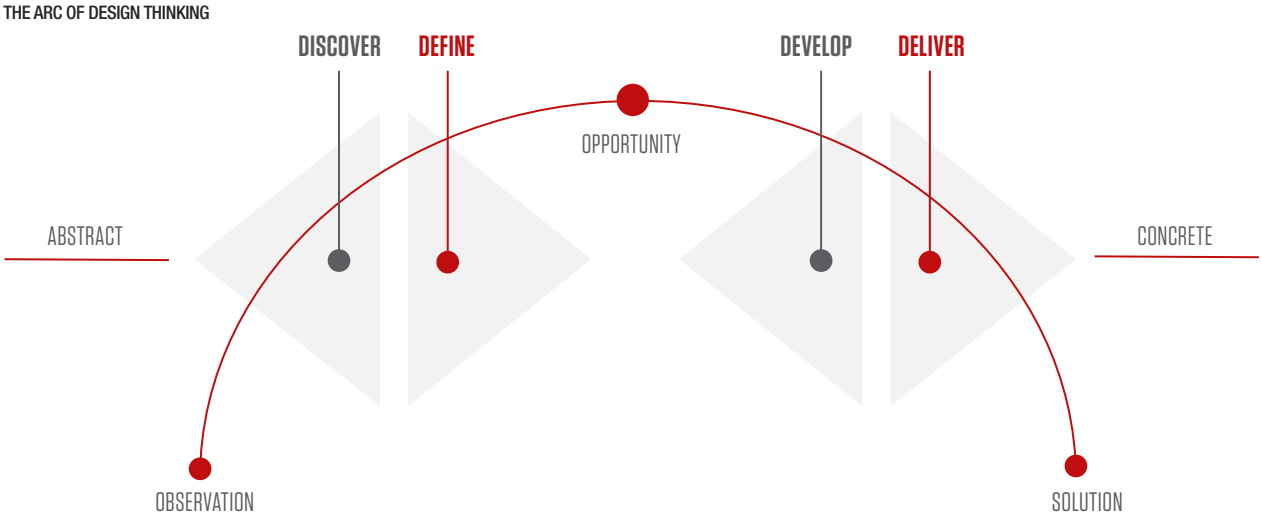
> "A key question for the financial service industry, as it struggles to reconnect with customers and staff amid a crisis of legitimacy and trust following the global financial crash, is how much impact design thinking might have in a field where its influence to date has been less evident than in other spheres."

Case study 1: Designing a hybrid bank branch network

In an era of digital disruption with more customers going online, a large Italian banking group wanted to rethink its approach to designing a future bank branch network so that it could draw customers in and connect with local communities as well as improve working conditions for employees and optimize the property portfolio. The project looked at the trend towards opening "hybrid" stores, combining bank branches with bookshops or coffee shops, for example, in order to provide a more local and individual experience. This kind of hybridization is often very practical in the sense that having multiple services on a site both draws more customers in and invites them to stay.

The project used design thinking methodology to undertake a study tour looking at hybrid spaces in London and Milan in the "discover" phase of the project. The precise requirements of the bank and its customers were defined as employees and customers were engaged in co-creation workshops. This led on to the development of a flexible architectural "kit of parts" to aid the converting and re-purposing of bank branches for more hybrid uses.

Figure 3: Diagrams showing how project stages combine convergent-divergent thinking with the concrete-abstract balance



Source: The Helen Hamlyn Centre for Design, Royal College of Art

Mini-hybrid, midi-hybrid, and maxi-hybrid models were developed. In the later stages of the project, a real-world pilot was built in Milan for evaluation and iteration.

Ultimately, this design thinking project will provide an insight into how the bank can adapt its physical and digital spaces to provide the bank branch network of the future, reaching out to local communities in a more meaningful way at a time when so much engagement and interaction is lost through online banking.

Case study 2: Rethinking the financial office environment

A large Scottish banking group wanted to improve working conditions, optimize space, and enhance employee engagement in its Edinburgh headquarters. Our research team began the project with in-depth user research inside three media organizations operating at different scales, from a creative agency with 60-70 staff working in a converted warehouse to a global communications company with 4,000 employees based on an out-of-town campus. These firms are early adopters of new ways of working.

Findings were used in the 'define' phase to create an architectural framework for office interiors. This framework demonstrated how workspace could be redesigned to be more socially engaging and dynamic by addressing four elements: programmable surfaces, circulation, large objects for way finding, and points of social interaction. This abstract thinking was developed in the world of ideas and brought back into the real world for implementation in the corporate headquarters building. The project using design thinking principles tested a fundamental design proposition to create more ownership, social capital, and social cohesion within a given financial service space.

The framework was subsequently developed into an online toolkit designed to enable a global furniture manufacturer, Herman Miller, to collaborate with its clients on workspace analysis and employee consultation when creating new people-centered work environments. Wider applicability was only achieved, however, through a process of direct design ethnographic engagement with a communities of office workers.

Case study 3: Banking without barriers for older people

This project worked with an age charity and a U.K. retail bank to explore the issues and challenges arising from the rise of internet banking, which has reduced branch footfall significantly and prompted a number of retail banks to review their branch footprint. While welcomed by many, these developments can prove problematic for some customers such as older people, 75-plus, who are

unfamiliar with online banking and those with physical and/or cognitive challenges who are most affected by reductions in traditional branch services. In this context, the empathic and people-centric focus of design thinking was highly relevant.

In the "discover" phase of the project, the research team investigated how customers who are affected by branch closures may need support to identify alternative ways to meet their banking needs, and those who consider digital banking may need some assistance to overcome their concerns. Throughout the project, extensive mixed-method design approach was adopted that included case studies, interviews, discussion groups, co-design sessions, spatial prototyping, and workshops co-authored with older customers.

New solutions were developed with the aim of improving access to financial services in a digital age, benefiting everyone in society regardless of age or digital ability. The study was brought back into the delivery stage with targeted proposals fed into the bank's design and implementation teams.

6. CONCLUSION

As these case studies demonstrate, whether building a community-based bank branch network of the future, rethinking headquarters office space, or removing barriers to banking for excluded groups of older, vulnerable, and disabled people, design thinking has something to offer in terms of addressing the challenges around financial transformation. At its most basic level, the financial services sector needs to get closer to its customers and employees to raise standards, performance, and, ultimately, profitability. The empathic, iterative, experimental, and evaluative qualities that design thinking brings to innovation can be practiced by non-designers, with elements of co-creation, and it can be argued that these are precisely what the era of digital disruption in the sector now requires.

REFERENCES

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 Collins

Calabretta, G., J. Montaña, and O. Iglesias, 2008, "A cross-cultural assessment of leading values in design-orientated companies," Cross-cultural Management: An International Journal 15:4, 379-398

Carlgren, L., M. Elmquist, and I. Rauth, 2016, "The challenges of using design thinking in industry – experiences from five large firms," Creative Innovation 25:3, 344-362

Chae, S., 2017, "Design thinking in action: changing the public service model," QLA Quarterly 22:3, 15-19

Cross, N., 2011, Design thinking, Bloomsbury Academic

Curedale, R., 2015, Design thinking: pocket guide, Design Community College Inc.

Dorst, K., 2015, Frame innovation: create new thinking by design, MIT Press

Fraser, H., 2012, Design works: how to tackle your toughest innovation challenges through business design, University of Toronto Press

Johansson-Skoldberg, U., J. Woodilla, and M. Centikaya, 2013, "Design thinking: past, present and possible futures," Creativity and Innovation Management 22:2, 121-146

Kleinsmann, M., R. Valkenburg, and J. Sluijs, 2017, "Capturing the value of design thinking in different innovation practices," International Journal of Design 11:2, 25-40

Luchs, M., 2016, "A brief introduction to design thinking," in Luchs, M., K. Swan, and A. Griffin (eds.), Design thinking: new product development essentials from PDMA, Wiley

Mootee, I., 2013, Design thinking for strategic innovation: what they can't teach you at business or design school, Wiley

Szczepanska, J., 2017, "Design thinking origin story plus some of the people who made it all happen," Medium, January 3, https://bit.ly/2taChUM

HOW DBS EMBRACED DATA-INFORMED DESIGN TO DELIVER A DIFFERENTIATED CUSTOMER EXPERIENCE

JURGEN MEERSCHAEGE | Head of Culture & Curriculum, DataFirst, DBS
PAUL COBBAN | Chief Data and Transformation Officer, DBS
MARK ENGLEHART EVANS | SVP, Head of Experience, DBS Transformation Group

ABSTRACT

Design thinking has led to tremendous improvements in business practices and customer service at a multitude of organizations. DBS Bank in Singapore initially followed the standard playbook for design thinking, though it adopted a '4D' framework of 'discover, define, develop, and deliver'. It soon strengthened the standard practices by implementing data-informed design thinking, which enabled staff to embed data into their design thinking practices and deliver an even better customer experience. The results are self-evident, as DBS Bank was recently named the Best Bank in the World by Global Finance. This article explains the processes DBS Bank went through on its design thinking journey, provides tangible examples of how it uses design thinking, and outlines how it is building on its success by embedding data into design thinking practices more deeply to deliver even better results.

1. INTRODUCTION

As it went about making significant improvements to its service levels over the past decade, DBS Bank also went through an evolution in its practices. It started by focusing on 'Asian Service' and developed an innovative framework called 'RED' that started a cultural shift within the organization. To drive change further, DBS soon embraced design thinking as a core methodology to make the shifts more effective. And, after finding that design thinking methodologies alone did not fully enable the process changes it aspired to, it developed data-informed design as an enhanced methodology that delivers even better results. It is still continuing the process evolution, by refining its practices even further so that all staff can fully leverage the benefits of the enhancements the bank is making.

2. THE GROWTH AND BENEFITS OF DESIGN THINKING

Fundamentally, design thinking is about applying the principles of design to the way people interact with the world. "Design thinking is a human-centered approach to innovation that draws from the designer's toolkit to integrate the needs of people, the possibilities of technology, and the requirements for business success," as Tim Brown, CEO of IDEO, explains it.[1] While design thinking may seem like a relatively new concept, it has actually been around since the 1960s.

[1] Brown, T., "Design thinking," https://bit.ly/2wRvxto

Design thinking uses an iterative process to understand the customer, challenge assumptions, and redefine problems in order to identify solutions that might not be readily apparent using other methods. It requires a deep understanding of the people for whom a company is designing products and services, through observing and understanding the target customer. Teams and individuals can use design thinking to combine what is needed by the customer with what is technologically feasible for the business.

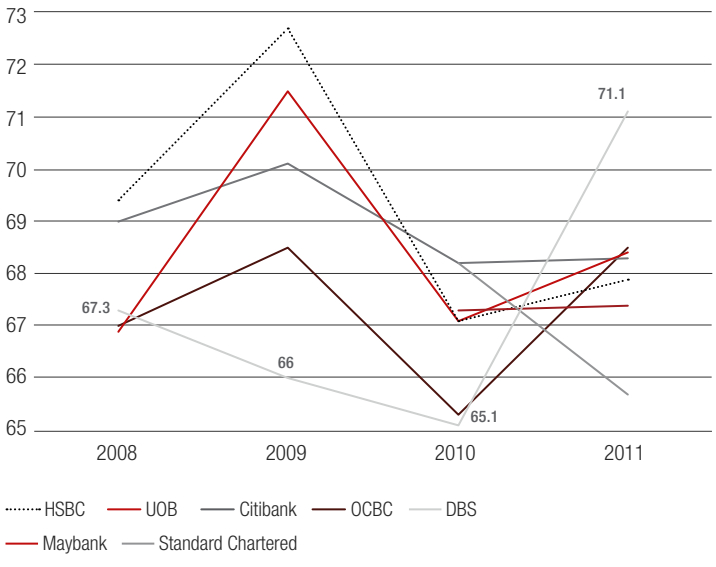
The process for design thinking most frequently has four or five steps, depending on the practices of the firm that is using it. IDEO, recognized as a leader in design thinking, uses a five-step approach that it says brings together what is desirable from a human point of view with what is feasible[2]:

Table 1:

APPROACH	ACTION
FRAME A QUESTION	Identify a driving question that inspires others to search for creative solutions
GATHER INSPIRATION	Inspire new thinking by discovering what people really need
GENERATE IDEAS	Push past obvious solutions to get to breakthrough ideas
MAKE IDEAS TANGIBLE	Build prototypes to learn how to make ideas better
TEST TO LEARN	Refine ideas by gathering feedback and experimenting forward

Source: DBS Bank

Figure 1: DBS customer satisfaction comparison with peers



Source: DBS Bank

The results of design thinking are impressive. Research by the Design Management Institute and Motiv Strategies that is incorporated into their Design Value Index (DVI), for instance, shows that design thinking-driven companies such as Apple, IBM, Nike, and SAP outperformed the S&P 500 by more than 200%.[3] "We see design not as a pure factor that makes our DVI company's stocks perform better on the stock market," said Motiv Strategies CEO Jeneanne Rae, "but rather as a highly integrated and influential force that enables the organization to achieve outsized results."

Banks have been relative latecomers to using design thinking, as they have concentrated more on maintaining current services for their large base of customers than on using design thinking to find out what customers actually experience and developing services to meet their real needs. More recently, however, leading banks have begun to use design thinking to transform their organizations. BBVA, for instance, has trained more than 1,000 design thinking experts to educate employees about how to apply design thinking throughout the organization. Capital One has established Capital One Labs as a laboratory within the bank to lead innovation using design thinking, and Lab members have opened 'Capital One 360 Cafes' as hybrid coffee shop-bank branches where Labs employees interview cafe customers to gather feedback on new prototypes.

Gradually, then, banks are beginning to use design thinking to innovate more effectively, by co-creating products with customers in order to integrate feedback more quickly. Banks are hiring teams with diverse

[2] IDEO, "Design thinking," https://bit.ly/2wRvxto
[3] Debbie Y., 2017, "The design value index shows what design thinking is worth," Fortune, August 31, https://for.tn/2OJAxfD

Figure 2: The RED service standards

Source: DBS Bank

backgrounds to build solutions with empathy, instead of using teams of people with similar backgrounds and strengths. When they build new products, staff in banks that use design thinking are also collaborating with other parts of the organization and with third parties instead of working independently in silos that do not communicate with each other.

3. USING DESIGN THINKING TO DELIVER ASIAN SERVICE

The journey that took DBS Bank in Singapore from being ranked last among its peers in customer service in 2009 to being named the Best Bank in the World by Global Finance in 2018 did not actually begin with design thinking. Over time, however, design thinking became the catalyst that propelled the bank into becoming a global leader.

3.1 Beginning a transformational shift

The transformation at DBS Bank started with an initiative that resulted in a bank-wide program called RED.

Coming out of the global financial crisis in 2009, newly-appointed DBS Bank CEO Piyush Gupta started working towards developing a strategy that differentiated the institution from western banks. DBS had come to the understanding that simply cleaning up the organization and making it more efficient was not the way to win. The leadership team soon coalesced around the concept of 'Asian Service' as the key differentiator, focusing on delivering a customer experience that would make banking joyful. Beyond just having a concept, however, DBS found that it had to determine what 'Asian Service' actually meant. After extensive analysis, it identified three core service standards, which it termed RED:

- Respectful to the customer
- Easy to deal with
- Dependable

DBS then put a process improvement program in place, leveraging lean techniques that would cut waste out of the system and deliver RED-level services to its customers. Within a relatively short period of time, DBS calculated that it had reduced the time customers expended on services at the bank by a total of more than 250 million hours. By 2011, DBS had moved from last to first in rankings of customer service at banks in Singapore.

3.2 Starting the design thinking journey

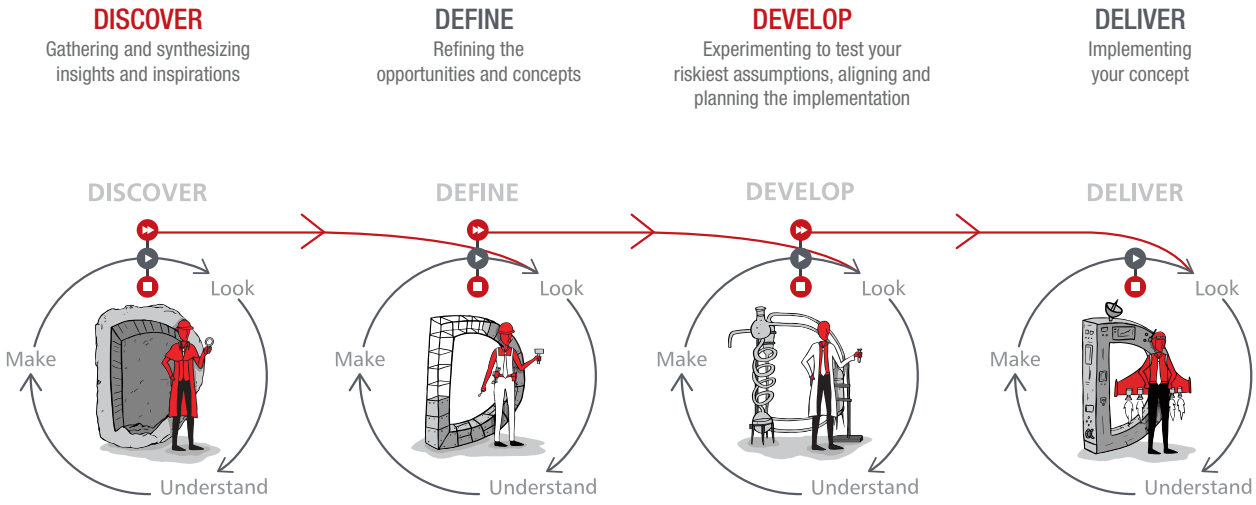
After implementing RED and improving service levels, DBS looked for a way to embed the core RED values into its culture. It soon identified design thinking as a possible way forward. One key insight in 2013 became the catalyst that pivoted DBS to embrace design thinking fully.

The bank was conducting process improvement events (PIE), one of which involved replacing lost credit cards. After reducing the timeframe for replacement cards from five days to one day, staff phoned a customer to request feedback. While the customer was pleased with receiving her replacement credit card quickly, she asked where her debit card was. She had lost her handbag and needed her debit card replaced too. It had not arrived.

The lightbulb switched on. Bank leaders realized that staff were too focused on processes inside the bank, rather than on the customer perspective, and were neither identifying nor resolving the full problem. Bank leaders tapped into staff who had expertise in design thinking and began the process of embedding it within the bank.

As it began to implement design thinking, CEO Piyush Gupta initiated a customer journey program by asking each senior leader to sponsor the use of design thinking to improve at least one customer journey. The journey program became a prominent part of the scorecard used to evaluate every managing director (MD). While not every journey was successfully transformed, pursuing large numbers of projects was essential to changing the culture, embedding design thinking into the DNA of the bank and delivering an amazing customer experience.

Figure 3: The 4Ds framework



Source: DBS Bank

To support the design thinking initiative, DBS established an innovation team with internal design thinking capabilities who have the capacities of an ad agency, service designers, and developers of user experience (UX) and developer experience (DX) strategies.

Gupta also decided, however, that he did not want this team of experts just to stay in a lab and coordinate the design thinking process. Instead, he wanted a team that would focus on problem-solving and enable staff throughout the bank to use design thinking tools in their jobs. A significant portion of the effort, then, revolved around education, coaching, and consulting. The bank subsequently reached the stage where most business units have the capacity for design thinking, even if staff from the core design thinking team are not participating formally.

To make design thinking more relevant to its specific needs, DBS adopted the '4D problem solving framework' from the British Design Council: 'discover, define, develop, and deliver'. 'Discover', for instance, involves understanding what a customer actually experiences. 'Define' requires identifying the problem to solve and determining whether it would be a worthwhile project, using feedback points from the 'discover' process. In the 'develop' phase, teams start designing products. Along with developing an actual product design, they also look at feedback loops that they can embed in the process so that they can measure the hypotheses, determine whether they are correct, and decide whether they are enabling the business impact they want to accomplish. Once they have verified that the product or service meets their goals, the teams 'deliver' it to the customer.

To marry design thinking with RED, DBS developed 'journey thinking' as its practice to use the 4D methodology to deliver a differentiated customer experience. During its design thinking journey, the bank found it essential to build cultural and business outcomes and to consider alternative ways of solving issues from a customer perspective or a design perspective. To ensure effective outcomes, DBS found that it is important to embed journey thinking into the DNA of the organization, focusing on building the right culture as well as providing superior customer experience outcomes. Journey thinking turned out to be easier for staff to embrace than originally expected, albeit still requiring significant effort and mindset changes, and it created confidence. The key message throughout the process has been that staff need to spend time to find insights to understand a customer's 'job to be done'. As Clayton Christenson explains it,[4] a businessperson who understands the job a customer needs to get done, designs a product to do that job, and delivers it in a way that reinforces its intended use will find that customers hire that product when they find themselves needing to get that job done.

[4] Christensen, C., 2004, "Marketing malpractice: the cause and the cure," Harvard Business Review, December, https://bit.ly/1P4qQmH

People who had one hypothesis often found, for example, that the customer's job to be done was actually different. The process, which leverages lean and agile, as well as innovation methodologies, aimed at a cultural shift rather than just having teams do things in isolation. Transformation has been broad and deep, allowing staff across the bank to participate and applying management resources to make sure the concepts were embedded deeply even into lower levels in the bank.

When they used design thinking to analyze the card replacement process, for example, staff found that the customer had three jobs to be done: I need to get home; I need to block my card; I need to get my life back in order. When it found it was not fully supporting the job customers needed to get done, DBS changed its call center practices. Agents first empathized with distressed customers who had lost their cards and then explained the replacement process fully. Contact center agents also offered to help the customer replace their other DBS cards and even assisted them with replacing cards from other banks by providing phone numbers for their contact centers. Customer satisfaction skyrocketed.

Since the journey thinking program began, the bank has used it for more than 500 journeys. The process is truly embedded in the company, and staff know that they have to map journeys thoroughly if they want to receive management support for developing any new products or services.

Very importantly, once it began to use and embed design thinking practices throughout the bank, customer satisfaction increased even further.

4. REAL TRANSFORMATION TRANSPIRED BY SHIFTING TO DATA-INFORMED DESIGN THINKING

While journey thinking clearly had a large positive impact on making banking joyful, DBS wanted to do more and make using data an essential part of its design thinking practices. The fundamental methodology for design thinking, the team observed, is traditionally qualitative. Staff put themselves in the customers' shoes, have empathy for the customer, and design products or services to meet customers' needs. Staff use interviews to discover customers' needs and understand what the customer is telling them.

As it focused on using data in every process within the bank, DBS started to encourage its employees to be more data-driven in their journey thinking. It also established a DataFirst team to turn DBS into a data-driven organization by focusing on developing a data-driven culture, data capabilities, data access, and data infrastructure. In the 'discover' phase of design thinking, for instance, DBS wanted staff to go beyond just putting themselves in customers' shoes, so they could fully understand what a customer goes through when dealing with the bank. DBS had followed traditional design thinking processes and taught employees to interview seven to ten customers. While that process provides useful information, it still may be anecdotal and not describe the full customer experience.

DBS thus began to teach staff to move beyond using information from the interviews and started having them also leverage data from far broader swathes of customers. Staff were able to identify customers' actual jobs to be done and the frictions they encountered, such as failed transactions or difficulties when they called the contact center. The result has been that staff using journey thinking can become familiar with customer needs quickly and use the tools as a source of data-informed design for the bank. It also turned out that data provided insights into whom to interview by identifying customers who actually had a problem or who were extreme users.

In the 'develop' phase, where DBS staff were designing products, data helped especially with two aspects. One was to enable staff to experiment at scale, with far larger numbers of customers. The second was to identify feedback loops at the end of the 'develop' phase, to measure whether their hypotheses were correct and whether they enabled the business impact that they expected. Whereas those data points are often an afterthought traditionally, DBS realized that staff needed to think about these questions upfront and embed a process for answering these questions into its products. Just like how companies such as Amazon, Uber, and Grab provide a personalized digital experience based on data, so too is DBS implementing similar feedback loops to get to the same level of personalization.

4.1 Overcoming misconceptions and challenges

DBS has also undertaken initiatives to overcome misperceptions or difficulties with using data.

One misconception is that some people believe they cannot be data-informed because they are not mathematicians. To overcome the issue, the DataFirst team emphasized that one of the most important aspects of being data-driven is the ability to identify the actual problems to solve and the ability to ask great questions that can be answered with data.

> "Just like how companies such as Amazon, Uber, and Grab provide a personalized digital experience based on data, so too is DBS implementing similar feedback loops to get to the same level of personalization."

Another shift that DBS has made to overcome difficulties is to use experimentation to validate conclusions from interviews and analysis. In the 'define' and 'develop' phases, for example, staff would interview at least eight customers. Originally, staff took positive customer feedback as validation for a new product. Over time, however, they found that responses from such a limited number of customers were insufficient. The bank moved to using data for digital experimentation, conducting AB testing, or testing through prototypes as well as other methodologies, which are broader-based and cost-effective. DBS has been able to lower the cost barriers so that it can test new products or services with hundreds or even thousands of people, enabling staff to know whether a product will have the traction they intend and expect.

4.2 Examples of the impact of data-informed design thinking

Examples from throughout the bank demonstrate the results that data-informed design has delivered.

As an example of data-informed Discovery, DBS wanted to go beyond improving the turnaround time for card replacement and eliminate a key friction point for customers by giving customers a digital way to request card replacements. Data analysis showed that the customers who most frequently requested card replacements were male students between the ages of 20 and 30 who frequented bars often. The analysis also showed that some customers had as many as 16 card replacements within six months. Staff then used the results of the analysis to identify the types of customers to interview and, using insights gained from talking with these two specific groups of customers, pinpointed customers' real pain points and went on to 'develop' solutions.

Another example, in the 'develop' phase, was credit card activation. When customers receive a new card, they are supposed to activate the card before they use it. Data showed that many new customers tried to use their cards without activating them, which resulted in transactions being declined and was often followed by a call to the contact center. An experiment was launched whereby SMS text messages were sent to a small percentage of customers when their transactions were declined, explaining what happened and how to activate their card. Analysis of this customer group showed that card activation rates improved while calls to the contact center reduced dramatically, and the practice was rolled out across the customer base.

Data on failed transactions has also been used to improve the customer experience. In India, for instance, DBS staff used data from the Digibank app and found that a significant portion of customers had unsuccessful logins because they forgot their password. Data alone could not explain the reason, so staff also interviewed customers. They found that the bank's password format, which was the same as in Singapore, was different from all the other banks in India. Once DBS changed its password format so that it conformed to Indian standards, unsuccessful logins dropped dramatically.

For business banking, DBS talked to treasurers and cash management staff at clients extensively to identify their needs. The design thinking team found that treasurers need analysis of information real-time, such as fluctuations in currency rates. The bank developed a scenario engine called DBS Treasury Prism to test moving cash between countries and the impact on the bottom line. It used customer data to find out how people are using the engine and where customer pain points are, then iterated alternative designs so that Prism could become easier to use.

5. CONTINUOUS IMPROVEMENT MAKES DATA-INFORMED DESIGN EVEN BETTER

As they analyzed the results of design thinking, bank management found two key results. One is that DBS is, indeed, making banking more joyful. The second is that the bank is better able to compete, whether with small fintechs or other banks or big techs, such as Google. While fintech competitors may offer an amazing customer experience, they typically focus on just one segment of financial services and do it extremely well rather than providing the full suite of financial services that DBS offers.

> "One misconception is that some people believe they cannot be data-informed because they are not mathematicians."

Even as the impact of data-informed design has been tremendously positive, DBS has continued to innovate so that it can deliver better service to its customers.

A key focus in 2018 has been to implement two new practices – 'instrumentation' and 'fulfilment' - in order to implement data-informed design even more effectively after products go live.

As it looked at what data to collect, DBS found that there were sometimes dichotomies between different teams. Whereas marketing and sales staff are typically more interested in successful transactions, for example, designers are more interested in unsuccessful transactions and other data points that would help them better understand customer friction.

In 'instrumentation', staff identify and develop feedback loops when they are building products so that they can measure and visualize the overall customer experience. Where capturing and analyzing data was traditionally an afterthought, instrumentation changes that mindset and embeds data user stories in the actual product design. Well instrumented products visualize customer friction points, which allows the team to continue to innovate and improve their products.

'Fulfilment' focuses on how to use data and adapt the experience in real-time to deliver a truly personalized experience. For instance, DBS customers will see personalized ATM menus based on their available account balance. This level of personalization requires deep instrumentation of all its products and services, and it is the most mature level of data-informed designed being taught in the organization.

As senior management reflected retrospectively on how to broaden the impact of the changes and looked at what has worked best, they also decided to focus more on culture by design. DBS has then worked to define its internal culture differently and change employees' mindsets by redefining its culture, describing the bank as a 26,000-person start-up with five key characteristics that it labeled ABCDE: Agile, Be a learning organization, Customer-obsessed, Data-informed, Experiment to take risk.

To embed data-informed design even more firmly into the bank's culture, DBS has used a curriculum around journey thinking and data. DBS has also put data-informed design into its coaching, clinics, workshops, and training. The bank publishes stories about what 'good' looks like, for example, and about how staff can run workshops themselves. DBS staff have learned to look at what data they will extract and how they will use it for ambitious innovation, differentiating the customer experience and next-level-productivity.

RED has evolved too, particularly as DBS has become more digital. Having digitized the back and middle offices, leaders wanted to know whether the bank was still being RED to its customers and following its core service standards. A key challenge recently, for example, has been defining what it means to be respectful to mobile-only consumers. While 'dependable' is more straightforward, with indicators such as server uptime and webpage load speed, 'respectful to the customer' is more difficult to define in a fully digital environment. DBS is in the midst of revitalizing RED in a corporate transformation project that revolves around defining what good service looks like for the digital future, making sure it is still following the RED mantra, and using digital power tools to achieve its goals.

6. CONCLUSION

Design thinking by itself has indeed led to many improvements at companies around the world. Corporations, consultants, and coaches have developed a variety of frameworks to make design thinking even better.

While qualitative practices have delivered excellent benefits, adding quantitative practices can add even more value. Some companies have struggled, however, to embed data into their design thinking processes. DBS has clearly overcome that hurdle, by using innovative techniques and teams to ensure that data is a core part of design thinking.

The results at DBS show that data makes the results of the design thinking methodology even better. It has used data-informed design to increase efficiency, reduce costs, vastly improve customer satisfaction, and innovate in ways that keep it ahead of even the fintechs that are developing sophisticated and focused products for consumers.

Examples throughout the organization, from retail banking and wealth management to corporate banking and human resources or marketing, have shown that data-informed design can work in every part of the bank.

The DBS model, and its process of continuous improvement, thus offers an example of how banks can use data to enhance design-thinking practices, so that they can compete more effectively amidst the fintech wave and increased competition that is inundating banking around the world.

EMPATHY AND CO-CREATION IN CAPITAL MARKETS OPERATIONS – INSIGHTS FROM THE FIELD

AMIR DOTAN | Principal Consultant, Capco Digital

ABSTRACT

Co-creation and empathy are fundamental principles of design thinking that enable teams to collaborate and solve user problems at pace. Cross-functional collaboration and deep understanding of end-users help to break down barriers between organization silos, resulting in an aligned vision and more holistic, user-centered solutions. However, the geographically-dispersed nature of investment banks can make co-creation and empathy-building challenging. Remote access to end-users makes it difficult to be immersed in their environment, and it is not always possible to instigate hands-on, face-to-face design workshops to foster co-creation. Leading design thinking projects in capital markets operations, we have observed that despite these constraints and limitations, the mindset and its application as a methodology to deliver solution definition can have significant benefits in an area of investment banking, which often lacks creativity, agility, and a user-centered mindset when defining strategic solutions. This paper provides an overview of the benefits and challenges of applying design thinking in capital markets operations, based on two case studies from Tier-1 investment banks, where we applied the approach to run solution-definition projects over a 12-week period. We share our experiences as design thinking practitioners and provide recommendations for achieving effective co-creation and empathy-building in a challenging work environment that has a lot to benefit from these two aspects of design thinking, and the mindset in general.

1. INTRODUCTION

Investment banking operations is a function that is constantly seeking to innovate and improve processes and operating models to meet key objectives such as regulatory compliance, reduction in operational risk incidents, and operational costs, to name just a few. Legacy processes, sub-optimal operating models, and ageing applications often result in inefficient, labor-intensive workflows that increase costs and risks. They are also costly to maintain and support, especially in a dynamic landscape that is affected by changing regulations and macro events like Brexit.

While some problems in operations have a clear and obvious solution – such as a fragmented data landscape that needs to be consolidated into a single source to be managed better - others may have the potential to be solved in a few different ways; sometimes without a technology component. However, in our experience "change management" initiatives often involve lengthy and detailed "current state" analysis and heavy technical documentation, which is not conducive to a holistic and creative process that explores a spectrum of options effectively to yield innovative, user-centered solutions.

In our experience, solutions are sometimes defined upfront, typically by technology stakeholders, with little to no consistent involvement of end-users in operations.

Figure 1: A common linear change management operating model in capital markets operations

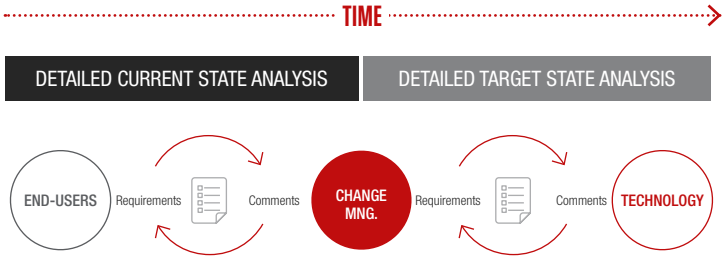
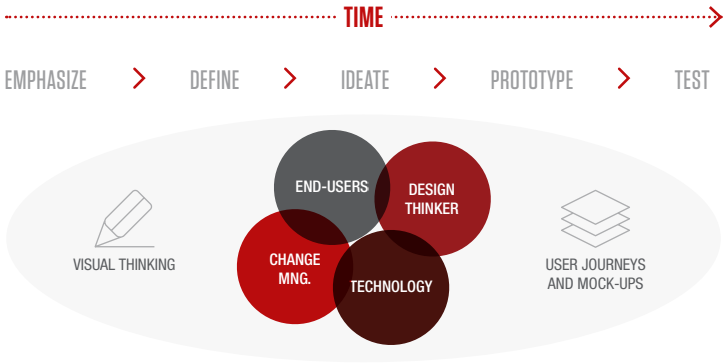


Figure 2: A design thinking approach brings stakeholders together to solve problems as a dedicated "solution team"



These could lead to solutions that are defined in isolation without a deep understanding and appreciation of the operational environment in which they are intended to be used to deliver value. A technical-led and highly-siloed approach to defining a solution, which we have consistently observed in Tier-1 operations, can limit the opportunity to effectively explore a range of creative ideas while leveraging diverse perspectives. As a result, the potential to innovate may not be fully realized.

It is in this context that design thinking, in our experience, can offer a more effective approach to innovation and strategic transformation, when tackling open-ended problems that have a strong human component. Applying design thinking to a solution-definition project can help break down barriers between internal silos to define - in a relatively short space of time - an agreed, long-term creative vision that is shaped very much with the end-users in mind throughout.

We have been engaged in several design thinking initiatives in Tier-1 operations in the last 18 months, where design thinking was applied to solve strategic problems in areas such as collateral management and cash settlements. Using the case studies we have worked on, we provide an insight into the day-to-day design thinking efforts in Tier-1 investment bank operations, focusing our description on empathy building and co-creation. For each we describe examples, challenges encountered, and lessons learned.

2. DESIGN THINKING – INTRODUCING A CULTURAL SHIFT AND NEW WAYS OF WORKING

While design thinking might conjure images of a group of people collaborating in a colorful space filled with post-it notes, user-journeys, and prototypes, the reality is that when working with multinational large organizations, such as a Tier-1 investment bank, the problem-solving process needs to be facilitated effectively across multiple locations - often remotely. Building empathy effectively with busy professionals around the globe, who execute highly complex processes, presents different challenges to engaging and immersing in the end-user's own environment.

The application of design thinking as a timeboxed innovation project within change management requires a different mode of operation than the one we typically encounter. As will be described in this paper, such an approach represents a considerable cultural shift that promotes strong cross-functional collaboration, iterative and creative approach with considerable focus on the human component of the problem and possible solution.

For someone who has not worked in an investment bank, such aspects may seem common practice and a prerequisite to driving the definition of innovative, user-centered solution. The reality is that introducing a mindset that champions co-creation and empathy within a multinational, complex, and often conservative organization can be challenging as it requires significant adaptation on the part of the firm.

As we will describe, getting stakeholders to adopt a mindset that can be radically different from their existing one poses logistical and cultural challenges. Injecting a design mindset into a risk-averse organization, and "change" function more specifically, is much more than simply introducing deliverables such as personas, user-journeys, and clickable prototypes. It fundamentally changes the way stakeholders engage with each

other and end-users for the duration of the project in order to accelerate an exploratory process to define a holistic solution.

3. THE CASE STUDIES

In this paper we describe two case studies where we have applied design thinking to solution-definition projects to solve strategic problems and deliver a long-term user-centered solution. As a mindset, design thinking can be applied in many different ways for different types of projects. Our focus is on projects within "change management" that are meant to produce long-term strategic solutions before the solutions can be implemented and delivered.

Case study 1 – collateral optimization

Optimizing collateral effectively is a top priority of "funding and liquidity management" to meet objectives such as cost reduction, as sub-optimal collateral is more expensive to deliver. While some aspects of collateral management are automated, others are manual and require the operations teams to carry out tasks such as contacting stakeholders via email or phone to initiate "collateral substitution" or execute collateral booking in response to receiving a margin call.

The initial, high-level business problem we were presented with was: "How can we optimize collateral without increasing operational headcount or creating unmanageable work for the current team?" Apart from the human component that needed to be acknowledged, understood, and considered, in the future collateral optimization solution, the optimization logic also needed to be envisaged and agreed by stakeholders from "technology," "operations," and "funding and liquidity management". Design thinking was chosen to bring people from relevant areas of the bank (operations, technology, change management, and corporate treasury) to shape a user-centered strategic solution for collateral optimization.

Case study 2 – equity swaps cash settlements

Equity swap is a highly bespoke and complex financial product, which can involve complex operational processes that require constant interaction between multiple internal teams and external stakeholders. Inefficiencies in the process and manual touchpoints contribute to operational risk incidents and reduced ability to ensure a smooth operation, whereby trades settle on the day they are expected to without any friction due to an array of possible problems. Troubleshooting such problems can be a time-consuming and challenging task depending on an individual's level of experience and familiarity with equity swaps and specific stakeholders.

The initial high-level business problem we were presented with was: "How can we improve the equity swaps cash settlement process in order to enable operations to better prioritize and manage their work, while reducing risk incidents and improving likelihood of settling trades on time?" A combination of factors, such as high attrition and the inability to effectively manage an increasing volume of work, meant that a long-term solution and a better understanding of the root causes of the problems experienced by operations were required.

4. EMPATHY

End-users are at the heart of the design thinking process as we aim to define solutions that will resonate with the target audience and address their problems effectively. The ability to put oneself in the end-users' shoes, relate to them, and feel their challenges, and not simply asking what they do and want, is a pivotal starting point. It ensures assumptions and bias are removed from the process in favor of uncovering people's real problems and context before possible ideas and solutions can be considered. Such level of understanding and appreciation requires effective immersion in people's environment to observe, learn, and gain deep insight into problems that need solving.

Abandoning assumptions in favor of "empirical thinking" (i.e., observe and learn) to build empathy is of particular importance in a professional, complex work environment like capital markets operations, where stakeholders outside the operations team often have very limited view and understanding of the nuances, idiosyncrasies, and challenging reality of day-to-day operations work. End-users can too easily be seen and defined narrowly by their role and responsibilities, e.g., "The collateral disputes analyst manages and resolves disputes on margin calls."

There are many qualitative aspects and layers to that person's role that are important to appreciate and recognize when aiming to solve a problem and define a solution that includes that person in some way. For example, the analyst works in a highly-social environment

Figure 3: People are a fundamental pillar of an operating model

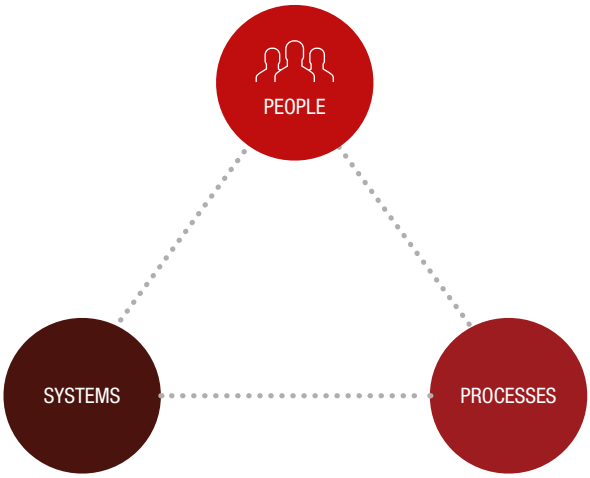
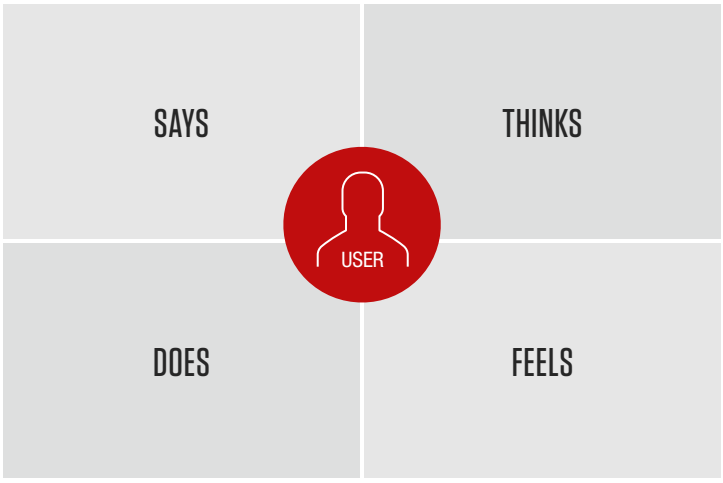


Figure 4: An "empathy map" is a useful tool to help stakeholders build empathy with end users



as part of a team of four, all sitting next to each other. As part of her day-to-day activities, the analyst spends an hour compiling end-of-day reports manually and running macros in Excel every 20 minutes to get an up-to-date view of prioritized disputes that need to be resolved.

Such actions are not only time-consuming and potentially risky from an operational perspective, but the low-level, mundane, and repetitive nature of such activities can be demoralizing. Being empathy-driven ensures that we do not lose sight of such important aspects when we define problems and explore solutions.

4.1 Empathy in enterprise environment

The need to start the problem-solving process with empathy is for the most part obvious when designing a consumer-facing proposition. User Experience (UX) and Service Design are well established concepts in personal banking, for example. However, when dealing with complex internal processes, especially in capital markets operations, terms like "empathy" and "user-centered design" are far from being as ubiquitous as they are in other parts of the financial services industry in our experience.

In such enterprise environments the end-users are a "captive audience". They are paid employees who are trained and tasked with executing tasks as part of wider operational processes. There is a constant drive by investment banks to improve operational processes by leveraging new technologies, for example "robotic process automation" (RPA) and "machine learning," which are expected to result in, among other benefits, freeing up people's capacity or making their role redundant altogether. In that sense, end-users can too easily be considered by some as a "moving part" in a complicated machine. The exact opposite of an empathetic perspective.

4.2 Starting with people

It is in this technology-led environment that design thinking introduces a fundamental cultural shift that places people as heart of the problem-solving process. The rationale being that people are a key component of a target operating model alongside systems and processes and, therefore, if people's real problems are not fully understood upfront, an expensive technological solution may fail to have the desired impact if it does not address real-user problems and challenges.

From what we have observed in the industry, typical change projects in capital markets operations tend to start with a technological solution in mind that is expected to deliver certain benefits, such as headcount reduction or increased throughput. End-users are contacted at the beginning of the project, typically by a business analyst, to provide input about how things are done currently, i.e., the "current state," as well as voice any requirements they may have for the future, to help form the "future state".

While this traditional "current state/future state" analysis approach may surface useful user requirements, it is not structured or executed as a user-centered exploratory process that focuses on generating insights rather than requirements. As a result, the output from such activities could lead to a narrow scope for innovation that is based on "what people say they do" and "what people say they want." In contrast, empathy makes it possible to develop a broader and richer understanding of people's situation and circumstances. As no solution is assumed at this early stage of the design-led problem-solving process, the empathy building activity is concerned with helping to generate useful insights to feed into the definition of the users' problems, rather than yield future requirements to a solution.

Listening to people's stories and experiences of their day-to-day work in operations - work that can often be confusing, stressful, and overwhelming - makes us not only informed but also emotionally invested. Empathy not only helps ensure stakeholders across the different functions have a clear understanding about what user-problems need to be solved and why, but they are also energized by the prospect of coming up with a good solution because they can relate and consider how different solutions may or may not fit the end-users' environment and solve their problems.

4.3 Empathize to break down barriers between silos

Despite the fact that an empathy-driven design thinking mindset is not common in capital markets operations, we have found that senior stakeholders were quick to acknowledge the importance of starting an innovation project with getting to know the end-users in order to keep an open mind, remove assumptions, and ensure that problems and context are thoroughly understood before potential solutions can be considered. This is a radical departure from the typical mode of operations, whereby a business analyst is tasked with creating detailed "current state" process flows diagrams as a starting point based on end-user input.

Moreover, in our opinion, enabling stakeholders from different areas of the bank to empathize with operations teams helped break down barriers between silos, by giving those teams faces, voices, and experience that made them far more than a functional role description. Investment banks are typically very siloed, both organizationally and psychologically. Empathizing with colleagues in different areas can be effective in softening the barriers between silos to enable more communication and collaboration when needed.

4.4 Case study 1

Approaching collateral optimization, it was key to get to know the work that was being performed by the operations team to understand the potential impact of optimizing collateral on an industrial scale. The engagement was designed as a 12-week process, with the first two weeks dedicated to "empathy building".

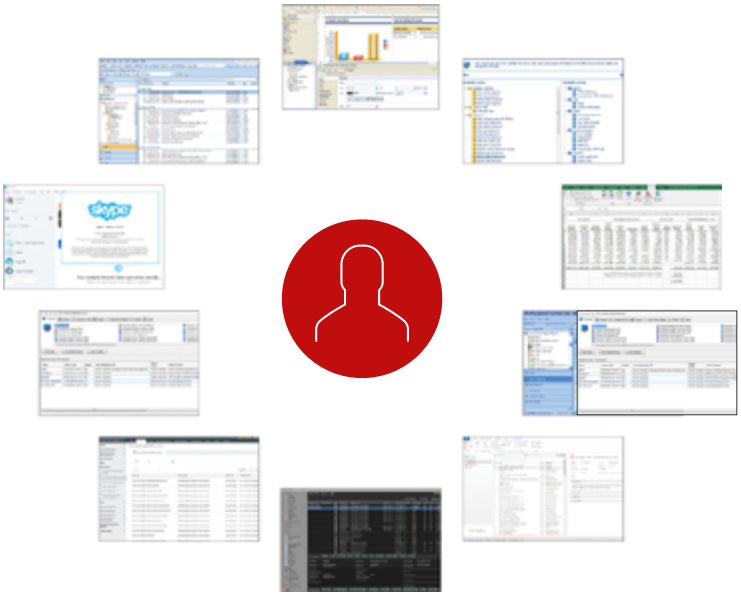
The aim was not to spend precious time documenting detailed workflows, but rather to form an impression of the work and identify key themes and areas that are most likely to be relevant to shaping a future collateral optimization solution. As a design thinker may not be a subject-matter expert, and given the limited time available, light-touch immersion to provide sufficient evidence and understanding of big issues (explicit and implicit) is all that can be realistically produced when dealing with a complex business domain in a short amount of time.

The relatively rapid nature of the design thinking engagement, compared to the more traditional analysis approach described earlier, meant that the initial discovery efforts had to be focused on useful and tangible outputs that could trigger empathy. During the 2-week period, we embedded ourselves within the team, implicitly observing the dynamic and initiating conversations with team members on an ad-hoc basis to clarify findings as we built our knowledge gradually. Sitting amongst the team, we were able to observe first-hand nuances, interactions, events, and conversations that are not likely to be captured in traditional business analysis methods.

Through daily conversations and observations, it quickly became clear the work was highly manual and attempts to optimize collateral on an industrial scale could likely increase this workload considerably, to the extent where it becomes unsustainable. For example, introducing a collateral optimization algorithm was expected to increase the number of requests for collateral substitution, which would need to be sent to external stakeholders by the team manually as things stood.

Whereas the team was typically tasked with processing 5-10 such requests for collateral substitution, it was plausible an optimization engine could initiate hundreds

Figure 5: Summarizing the end-users' application landscape in a simple visual way helps to empathize with the complexity and realities of day-to-day operations work



of requests, depending on how it was configured. The team's tasks were labor-intensive and members relied heavily on emails to communicate with internal and external stakeholders, alongside an array of disjointed applications. Any significant increase in the volume of items that would need to be processed daily would result in a major operational bottleneck.

Getting stakeholders to empathize with the end-users was achieved in part by highlighting some of our findings and sketching high-level user-journeys that communicated the nature and essence of the work that were important to appreciate in the context of possible future collateral optimization.

Being curious is key to effective empathy building in our opinion and looking around the end-user's environment can be very informative. For example, we observed that calculators were used during various workflows, whether physical or digital ones. This anecdotal finding brought to life very effectively the day-to-day manual work we observed, which was not familiar to the stakeholders outside the team. It served as a powerful reminder of some of the operational challenges that needed to be addressed in order to successfully optimize collateral on a grand scale.

Findings such as physical calculators not only made the work of the team feel very real for people that were not familiar with the work involved, but also emphasized the considerable gap between the sophistication the collateral optimizer represented, and the current manual operational work that would need to be considered in the overall solution. It left a lasting impression throughout the engagement - even in later stages of the project, senior stakeholders referred to it as an effective reminder of the end users' context. It is worth noting that taking photos in an investment bank has to be approved beforehand due to data security concerns. While it can be very impactful, it is not straightforward like it may be in other industries to video and photograph the end-user's environment.

4.5 Case study 2

Compared to the collateral optimization project, redefining an equity swaps cash settlement process to address strategic problems presented a different, more complex challenge for building empathy:

- **Multiple locations:** whereas the operations team in collateral optimization was based in London, the end-users in this case were in several locations around the globe. This meant that getting to same level of immersion and frequent contact would be impossible without traveling and spending time in each location.
- **Domain complexity:** equity swaps are highly complex and bespoke financial products with many moving parts. As a result, the operational process was significantly convoluted, involving many scenarios and workflows.

Building empathy with a global team under such circumstances may seem like a daunting task, especially given the 2-week timeframe to conduct the empathy building activities. We chose to conduct remote interviews with screen sharing to avoid losing time due to travel. Budget was also a consideration. While being physically present in the end-users' environment can be very insightful, given the number of locations traveling was deemed undesirable. In a sense, having limited contact with the end-users in this case (typically an hour interview and a shorter follow-up conversation) allowed us to focus on the high-value issues and aspects of the complicated work rather than getting caught up in low-level, tactical details.

Through the conversations with the operations team, a picture emerged of a situation in which work is carried out in "near-zero visibility conditions" as team managers were often pulled into work that took them away from managing, putting them in the dark about what team members are working on and why. As a result, they struggled to prioritize, plan, and manage effectively. Screensharing during the interviews was essential to bring to life the highly fragmented work environment, showing us the multiple systems involved in workflows that are often not documented clearly and need to be "learned on the job".

4.6 Recommendations for empathy-building

As described above, empathy can be very powerful to ensure that efforts to innovate focus on the right areas and that solutions are designed with the end-users' environment, context, and needs in mind. When approaching empathy in capital markets operations we recommend the following:

- **Look for stories, not just requirements:** when meeting end-users and talking to them about what they do, it may be tempting to note down in detail every step of every process that is being executed. Instead, empathy is more likely to occur when you approach the conversation with the aim of capturing the end-user's stories and anecdotes. These can be very powerful to convey the essence of the job and the big problems individuals face on a day-to-day basis. It requires good active listening skills and genuine curiosity to ask insightful and follow-up questions. You know you are successful in your efforts when you feel passionately you want to solve the end-user's problems because you understand the impact they are having.
- **Create impactful visuals:** conveying to others what you have discovered about end-users so that they can also empathize can be challenging, which is why creating impactful visuals (e.g., photos, illustrations) can be very effective to tell the story and bring findings to life. An audience is far more likely to take interest when presented with engaging content, such as high-level user journeys, which are designed to describe to people the essence of what is being done, rather than attempt to document every step and decision point along a complicated workflow.
- **Use end-users' time wisely:** the people we often aim to empathize with are very busy and are engaged in time-critical tasks. We do not have the luxury of spending a lot of time with them, nor can we realistically get our heads around everything they do. All this means that we need to ensure the time we do spend with them is very effective. We often achieve that by focusing on fairly high level "stories," such as a typical day that provide enough detail about the day's "highlights" and key problem areas. This can be achieved in half an hour if done right. The relatively short amount of time available means we need to get to core issues quickly, and get a summarized view that can lead to useful insights.

5. CO-CREATION

Design is very much a cross-functional activity that relies on combining diverse range of views, experiences, and perspectives to craft a solution. Co-creation - the coming together of stakeholders from different areas to explore and define a solution collaboratively - is a fundamental pillar that underpins design thinking. Collaboration not only allows for different perspectives and ideas to be shared quickly, ensuring ideas are acceptable to all they concern, but in a large organization it can be highly effective to save time and break down barriers between silos as stakeholders from different areas of the bank come together, often for the first time, to collaborate in co-creation design sessions.

For us, co-creation is a hands-on approach, using tangible design artefacts, such as storyboards and prototypes, to quickly visualize ideas so that they are easier to envisage and understand as they take shape and are iterated upon quickly. This is in stark contrast to the common mode of operation we often come across in capital markets operations and technology, in which one group of stakeholders produces or owns the content – typically a business requirement document (BRD) - and periodically shares it with a wider audience to be discussed over a meeting in a boardroom-type setting.

5.1 Challenges to co-creation in capital markets operations

In our experience, introducing co-creation as a way of working in capital markets operations can prove challenging for two main reasons:

- **Logistical:** applying a hands-on approach where people come together in a shared physical space to create artefacts, such as mock-ups and user-

journeys, is challenging when teams are spread across the globe, as is often the case in global investment banks. Remote collaboration tools, like Skype and video conferencing, have their benefits, and offer a useful alternative to not collaborating at all. However, they also have their limitations when considering such hands-on activities. It is also worth noting that securing a decent sized meeting room in a bank is not always easy, as such spaces are very much in demand. Meeting rooms do not always have whiteboards and we have been in situations where it was not allowed to post things on the walls.

> "Applying design thinking to a solution-definition project can help break down barriers between internal silos to define – in a relatively short space of time – an agreed, long-term creative vision that is shaped very much with the end-users in mind throughout."

- **Cultural:** stakeholders in our experience are often comfortable attending meetings remotely using Skype, even if a meeting room is available. It may be because it allows them to multi-task in the background, and we have encountered situations where people needed considerable persuasion to leave their desk for a couple of hours and join a face-to-face co-creation session. Additionally, oftentimes people consider themselves consumers of content produced by someone else rather than co-creators of the content itself - especially if the other potential co-producers are from a different area of the bank. As such, transitioning from consumers to co-creators represents a considerable paradigm shift for some.

5.2 Mobilizing a co-creation 'solution team'

To facilitate frequent, effective co-creation during a design thinking engagement, we assemble a dedicated internal group of 10-12 persons we call a solution team. It is tasked with coming up with the solution and at the very least includes stakeholders from technology, change management, and end-users from operations. Team members are expected to commit up to six hours a week to participate in co-creation workshops, either in person or via video conferencing. This is a crucial prerequisite to effective design thinking, and while it may seem a considerable time commitment to some, we have found it ends up saving far more hours of deliberations and content creation in the long term.

As design thinking practitioners, we work with the solution team, carrying out empathy-building activities, guiding the design thinking process, facilitating the co-creation sessions, and producing the project output and design artefacts. We have found that having such a solution team makes it easier to promote diverse thinking and tap into internal creativity, which people may not feel they have the opportunity and license to express otherwise. It also creates a sense of shared ownership and ensures stakeholders from different areas have 'skin in the game'.

Keeping the size of the team relatively small is key to ensuring that team members can collaborate frequently and effectively. This is a very different approach to a typically large "working group" scenario we have encountered on many occasions, which can turn counterproductive. We often see change projects involving large groups of up to 20-30 individuals who try to resolve issues and come up with solutions during weekly conference calls. This is a very common mode of operation in capital markets, especially since stakeholders are often located in multiple locations.

To ensure the solution team has first-hand knowledge of the day-to-day reality of operations work when considering user problems, ideas, and possible solutions, we ask that members include junior stakeholders that perform the low-value manual work and not just managers. Such a request is often initially met with raised eyebrows for two main reasons:

- **Organizational culture:** in our experience, it is rare for junior resources in an analyst/associate role to attend regular strategy meetings with managers at director, vice president, and executive director levels. A hierarchical corporate culture means an analyst in operations is removed from managers, except a direct line manager, and has little to no interaction with managers in change management and technology. Given that the analyst is often a primary end-user, who very likely will be part of an operational solution, their involvement as a member of the solution team has

always proved extremely important to provide a unique perspective "from the trenches".

- **Use of time:** having analysts participate in regular solution team co-creation workshops meant they were not at their desk for one to two hours to respond to emails, phones, and execute their day-to-day tasks. While that is a legitimate immediate-term concern, it is important to recognize that the far more strategic and long-term solution-definition would likely suffer without constant input from the end-users. As a compromise, we often propose that end-users alternate, so that we get a mix of perspectives from the team, but also so that it is not always the same person that is absent.

5.3 Case study 1 – single location

Introducing co-creation as a way to solve collateral optimization was relatively straightforward from a logistical perspective, since all of the stakeholders were based in London, albeit in two buildings which were five-minute walk apart. This posed a slight challenge at the beginning when one key stakeholder expressed preference to attend the co-creation workshops remotely, despite the short walking distance. We felt that it was important that everyone in the solution team who was not working from home on the day of a workshop attends in person due to the interactive nature of the workshops, and to also allow the team to gel as a unit. The issue was resolved successfully and the fact that the solution team attended the sessions in person proved very beneficial.

Despite the fact that stakeholders from change management and the business (end-users) were sitting in close proximity, the solution team sessions were the first time analysts engaged with managers from change management and technology. Typically, a business analyst from change management would engage end-users to elicit requirements and would then document the material and present it to the project manager and change lead responsible for the business function.

The process of debating and co-creating ideas as a group was very efficient and highly productive for everyone to understand what role the operations team might play in the future to help facilitate collateral optimization without having a detrimental impact on day-to-day operations. The co-creation sessions took place in the client's corporate training facilities, which were ideal as they were large rooms with furniture that could be rearranged easily as well as large whiteboards. Low-fidelity storyboards, depicting in a highly-visual manner the future vision as a step-by-step narrative, brought to life ideas such as automated client emails, workflow

tooling, and straight-through-processing (STP) to free-up capacity for the team to handle high-value work that required human involvement.

> "Co-creation and empathy building are fundamental aspects of design thinking, which are crucial to explore and define creative solutions effectively and address real user-problems successfully."

Having representatives from technology and end-users working together was extremely useful and made it possible to refine ideas quickly while validating their technical feasibility and assess the extent to which they would be welcomed by the end-users. It was clear that technology stakeholders who had a deep theoretical knowledge of the business domains, learnt a lot in the process about what takes place in operations.

In some cases, stakeholders from technology and funding and liquidity management came to realize during the sessions that certain ideas and approaches that made sense in theory would not work in the real world from an operational perspective. For example, the bilateral nature of the interaction with individual external stakeholders meant that each stakeholder the operations team interfaced with posed different challenges and may require a different approach. Some were notoriously slow to respond, while others did not tend to reply to emails so phones were the main mode of communication. Assuming such behavior was not likely to change, it had to be factored into the solution in some way. We doubt such insights would have been acknowledged and considered if it was not for co-creation.

THE OUTCOME

In the end, the solution that was achieved represented the views, ideas, and input of all the stakeholders and each felt they owned it equally. We were able to outline an agreed solution that included a logic for optimizing collateral as well as new tools that operations would need to be equipped with to handle new high-value tasks as part of the long-term vision. This was a radical departure from the normal way change management had approached projects and there was a sense that consensus was established firmly, and everyone shared the same vision.

5.4 Case study 2 – multiple locations

Reimagining a back-to-front equity swaps settlement process proved far more challenging, the principle reason being that the stakeholders were spread around the globe. It quickly became clear that we would not have the luxury of face-to-face co-creation sessions in a shared physical space.

Gathering members of the solution team in the same space, even if only once or twice, was not practical for several reasons, and even if it was possible, the intense nature of the engagement over a 12-week period meant that it was not a solution for the entire period. We needed to come up with a strategy that could ensure productive and sustained co-creation between stakeholders in five locations in the U.K. and North America.

CO-CREATING REMOTELY

The solution was to mandate that all the co-creation sessions will be held via video conferencing to ensure there was a strong sense of presence in each location and ability to share, in real-time, the highly visual output that was produced in each location, often in a non-digital form (e.g., pencil sketches and storyboards).

The investment bank we worked with did not have access to advanced remote collaboration tools that may have made this process easy for us, and for obvious reasons, the process of authorizing and installing applications in such organization can be lengthy and success is not guaranteed – especially across multiple locations, which added an extra layer of complexity and risk.

Considering the logistical constraints, it was decided to go ahead and make the best of the firm's video conferencing facilities. Co-ordinating the booking of five meeting rooms with video conferencing capability created a constant administrative challenge, which we overcame for the most part by scheduling all the sessions well in advance.

THE OUTCOME

In total, the solution team spent 36 hours co-creating remotely from defining the end-users' problems following an empathy building phase through to testing prototypes at the advanced stages of the engagement.

Whiteboarding sessions in which user-journeys and ideas were brought to life through elaborate, rich storyboards required constant camera work, zooming in and out and panning left and right. At times it may have been slightly disorienting to the people in the different locations, but overall proved effective and the sessions' output did not suffer as a result.

Solution team members commented that leaving their workstations for two hours every few days to collaborate face-to-face, technically speaking, was refreshing and a highly effective use of their time. The highly visual nature of the activities and the tangible output contributed to high level of engagement from all involved. We felt that the co-creation aspect of the engagement encouraged people to think big and fast, as we were collectively exploring possible long-term solutions, rather than short-term tactical fixes.

The cross-functional nature of the solution team proved to be powerful and resulted in many interesting and productive discussions, primarily between end-users and technology stakeholders. As we often witnessed in such settings, there is usually a considerable gap between what technology and change stakeholders know about the work that is being done by the operations team and the day-to-day reality and nuances the work entails. The co-creation sessions in this sense are a great way for stakeholders outside the operations team to build empathy and establish deep understanding of the needs, circumstances, and user-problems that are pivotal to drive effective ideation and prototyping sessions.

5.5 Recommendations for effective co-creation

As has been described, facilitating effective co-creation is not always straightforward, as it is often a new mindset in capital markets, which requires people to work in a different way than they are used to. To make the process of co-creation effective and sustainable during the lifecycle of a solution-definition project we recommend the following:

- **Aim for in-person or video-enabled face-to-face sessions:** we cannot stress enough how effective it was bringing people together in the same room, physically or virtually, to co-create. We believe that the interactive, hands-on nature of co-creation design sessions makes it impractical to consider remote solutions that do not create a sense of real presence. Co-creation is more than having people talking and looking at the same PowerPoint presentation. Such mode of operation may be sufficient for quick update calls or ad-hoc reviews of emerging solutions, but to drive an effective exploration and co-creation people need to feel that the session is different from other conference calls they perform during the day. Co-creation should be engaging, energizing, and productive over a course of 8-12 weeks in some cases. In our experience, maintaining high levels of engagement is far more likely when people are expected to show up rather than dial in.
- **Keep the size of the solution team small:** there is typically in an investment bank a desire to get everyone involved in defining a solution, which can easily result in an unmanageable and counterproductive group of people. Because people are used to dialing into "mega" conference calls, this may not seem an issue, but when applying design thinking and wanting to get people active and engaged, ideally in the same space, a large group is not practical. Providing visibility to a wider group of stakeholders on a weekly basis as part of "working group" is one way we are able to keep the size of the solution team, who meets every few days for a couple of hours, small.
- **Arrange co-creation sessions well in advance:** the two main logistical obstacles we faced when attempting to introduce co-creation in an investment bank were finding meeting rooms and slots that everyone in the solution team could accommodate, at times across multiple time-zones. To address these issues, we book the sessions weeks in advance – basically, as soon as we know who the members of the solution team are. Changes to people's availability are inevitable and we constantly need to adjust and shift sessions around, but setting the meeting far in advance gives members of the solution team a good idea about the expected time commitment and overall project structure.

6. CONCLUSION

Co-creation and empathy building are fundamental aspects of design thinking, which are crucial to explore and define creative solutions effectively and address real user-problems successfully. As this paper described, both are hugely relevant to solve problems in capital markets operations. However, applying them in practice in large multinational organizations can be challenging for logistical and cultural reasons, to name a few potential obstacles that need to be overcome.

Barriers between organizational silos can easily get in the way of defining innovative solutions at pace that are fit-for-purpose. They can result in prolonged review cycles of solutions that are slowly being defined during conference calls, using heavy technical documentation that can limit, if not stifle, people's creativity, and ability to appreciate the user-problems that need solving. As organizations come to consider ways to encourage people in different business areas to collaborate more closely, and focus on real user-problems in the process, design thinking offers a structure and mindset that has co-creation and empathy building at its core.

As was described in the two case studies, ensuring co-creation and empathy building during a design thinking engagement in capital markets operations is not straightforward, and requires logistical preparations as well as winning the hearts and minds of stakeholders who are expected to adopt a different way of working to solve a problem. The benefits of early and continuous cross-functional collaboration become clear quickly, as people comment that they feel their time is being spent better and that having multiple perspectives in the room accelerates and improves the output, which is considered from the point of view of real people with real problems.

HOW DESIGN THINKING IS POWERING PAYMENTS INNOVATION: OUR JOURNEY AT MASTERCARD

KAREN PASCOE | SVP, Experience Design, Mastercard

ABSTRACT

Most companies are looking for ways to improve innovation and reduce the risks associated with development efforts. A key way for organizations to accomplish that is through the use of design thinking, which is a structured approach for cross-functional teams to develop and maintain a clear focus on user needs from conceptualizing an experience through iteratively prototyping and testing with the intended audience.

Design thinking has taken off in many organizations, from technology companies to corporates to non-profits. A key benefit from the approach is to break through inertia, build and maintain user empathy, and harness the comprehensive perspectives of a broader team.

Organizations who embrace these methods tend to find heightened employee engagement, better collaboration, and more innovative solutions. Organizations looking to embrace and scale design thinking can learn from Mastercard's journey from innovation lab to enterprise.

Design thinking is part of almost everything we do at Mastercard. We use design thinking to inspire research and development, as well as consumer product development. Design thinking helps us solve challenges with our customers and partners. We even use design thinking to improve our own employee experience.

A revolution is underway. The connection of the physical and digital world is transforming our industry (and many others). In this new world, every device can be connected and every device can be a commerce device. Consumers have extremely high expectations of the user experience and unprecedented choice.

Our business provides ways for our customers and partners to reach consumers with capabilities that safely, seamlessly power their lives in commerce no matter where they are. While we at Mastercard have always enabled consumer experiences, the experience standards of the digital world are changing the game for the broader commerce ecosystem. In response to that, we have activated design thinking across many areas of our organization to elevate consumer centricity and innovation.

1. SEEDING

Mastercard Labs, our research and development arm, has been using design thinking across our innovation portfolio for some time. We blended the best of formalized design thinking practices with lean/agile methods to accelerate progress. These techniques allow the cross-functional product team to focus on the consumer pain points that they are solving for. Design thinking is part of

the day-to-day activities of our innovation teams in our global hubs in New York City, Miami, St. Louis, Silicon Valley, Dublin, Nairobi, Singapore, and Sydney.

In Nairobi, Mastercard Labs for Financial Inclusion is focused on payment solutions as a way to positively impact low income households in Africa and other developing markets. Our belief is that bringing formal financial services will empower people and drive economic growth. We are also supported in this endeavor by the Bill and Melinda Gates Foundation.

A recent innovation out of Nairobi is 2kuze, which means 'let's grow together' in Swahili. In order to meet the needs of small scale farmers, our team worked with the farmers side by side to understand their needs and pain points firsthand. We took a prototyping approach to test and learn, working iteratively toward a solution that allows farmers to streamline sales processes, expand their pool of buyers, provide transparent pricing, and accept payments – in this case digital – without ever leaving their farm. Being in the field makes it easy for the product team to design these products with the actual users. The close proximity keeps a clear and continuous focus on user needs, resulting in better product experiences overall.

As a company, Mastercard is focused on doing well by doing good, so we have brought design thinking approaches to our work with humanitarian organizations. After extensive partnership with the Mastercard Aid Network, our design thinking approaches have allowed us to better understand the needs of everyone involved.

In this case, we worked to meet the needs of the population (which range from refugees to civil unrest to natural disasters), the aid organization, local merchants, and the donors who support these efforts. The solution we developed can be quickly implemented anywhere using a web-based solution for merchants and providers that can be operated in remote locations without connectivity. This cost-effective solution provides transparency to aid providers, streamlines the merchant experience, and preserves dignity for aid recipients in a time of great duress.

Figure 1: Karen (second from left) and her team iterating some design concepts in digital payments

2. MAINSTREAMING

Within our product development teams, we leverage a variety of different techniques including design thinking, customer journey mapping, lean UX, and a set of advanced human centered design methods where we have dedicated designers. As with many organizations embracing the digital world, we are in a broader transformation that is elevating customer experience, design, and innovation. This aligns with our technology modernization, incorporating APIs, dev ops, and microservices with agile approaches.

> "We benefit from blending different thinking styles that encourage teams to think bigger, suspending preconceived notions of solutions before they go deep, and edit their choices."

We are evolving as an organization to start with the desired experience and working through how to deliver that. Design thinking approaches allow the team to see the holistic experience through a consumer lens, which drives better 'outside-in' perspectives. We also do research and testing with consumers that keeps our teams focused on the right pain points all the way through our development cycle.

With customers and partners, we often have comprehensive relationships, complex implementations, and multi-year service periods for our programs. Design thinking helps us collectively define the consumer experience. We assemble the key participants, with the right end-to-end representation from both teams. Together we work at solving consumer pain points in a way that is highly engaging for all involved, which gets to better results than a traditional RFP process. Through rapid prototyping and testing, we can quickly and inexpensively get to consumer validated outcomes that reduce risk and improve overall results.

In some instances, we have customers who are looking for our help with a particular problem or segment. Many of our customers, for example, are looking to improve engagement with millennials or centennials (Generation Z). This is an area where design thinking shines. The team gets a clear consumer research activity and briefing, and they are given a set of activities designed to get them into the target mindset. Traditional methods simply cannot generate this level of consumer empathy.

All of this great work has not gone unnoticed by our customers and partners. By popular demand, we have created an offering from our consulting arm called Mastercard Labs as a Service (LaaS). Our teams are skilled in the facilitation of our agile design thinking methods, with backgrounds in design, research, innovation, and product development. They work closely with our customers to understand objectives and carefully curate an experience for the customer with their cross-functional employees. It is a highly engaging method that helps our customers drive employee creativity, supports their own digital transformation agenda, and provides highly innovative, consumer centric solutions. We drive to viable offerings with strong business cases that have scale potential.

The majority of our engagements are a five-day rapid design and prototyping service. There is significant definition and preparation in advance so that, from the moment the customer's team walks in the room, it is a high impact experience. Our offerings also extend to design and consumer testing through pilot development and product buildout. We also have innovation capabilities that we use via our own Mastercard Labs that we can establish for our customers to syndicate and scale proven techniques.

We are starting to use design thinking for our internal efforts as well. There is a focus on enhancing the employee experience, which is a foundational element of delivering a good customer experience. Using the same methods that we do for consumers, we have identified the key journeys for our employees, started prototyping experiences, and are working iteratively to get the pieces into play. All of this hand-in-hand with our HR teams doing the work. This was an excellent opportunity for HR leadership to see for themselves how working cross-functionally and starting with research, empathy, and needs gets us to better solutions. The team recognized they had more creative solutions to problems than expected and felt more engaged in problem solving.

Figure 2: Technology modernization into RESTful APIs, microservices, and agile practices will drive scaled innovation and better consumer centricity

3. TRAJECTORY

There are clearly a lot of applications for design thinking at Mastercard. I see us as being well along in our broader digital transformation journey. Lean, innovation, agile, design, customer experience, dev ops – all of these methods work to connect organizations – drive more creative problem solving and keep the focus on customer needs. We have been using design thinking in pockets for over five years. As these pockets have yielded great results, there is more interest. Our hiring continues to scale in user experience design, customer experience, and related fields; so our deep practitioner base continues to scale.

Traditional business and technology roles are undergoing a shift as we recruit more and more sales, product, and technologists who have experience working in customer immersed teams that leverage design thinking methods. Our technology teams are starting to speak in the language of customer journeys and asking richer, more empathic questions about our user base overall.

From a learning and scaling perspective, we have formalized a design thinking curriculum at two levels, an introductory course that has trained over 600 employees to date, as well as an in-depth facilitator training and certification program. Those completing the program are asked to lead design thinking facilitation sessions outside of their existing team at least quarterly. This helps us expand our pool of facilitators, keeping the knowledge base, and fostering a community of practitioners.

Our efforts are further amplified by a broader product development lifecycle refresh. This refresh, deeply informed by design thinking, is being rolled out to our employee base globally. We have done a benchmark of our process enhancements and find it akin to what we are seeing at leading innovators across a variety of industries.

These efforts are helping us scale innovation more broadly. As Mastercard and the payments industry are broadly transforming, we need thinking that keeps us highly focused on end-user needs. We benefit from blending different thinking styles that encourage teams

to think bigger, suspending preconceived notions of solutions before they go deep, and edit their choices. The clear focus on prototyping and iterating allows us to get input incrementally, reducing our risk overall. This helps us move beyond our traditional payment methods (physical cards) and evolve more deeply into digital and payments between businesses.

Historically, our traditional customers have used mature ISO connections into our payment processing systems to power payments to the over 2.3 billion consumer accounts. We have well over 23,000 customers who connect to us for settlement and our payment brand is accepted at over 45 million merchant locations. As we modernize our technology to RESTful APIs, microservices architecture, and platform thinking, we create new options for innovation. This will create a virtuous cycle of innovation as we continue to evolve who we partner with, the markets we serve, the types of payment capabilities we provide, and the ways that payments are evolving into IoT and devices more broadly.

4. CONSIDERATIONS

Design thinking helps us look beyond our own internal groupthink to get to new ideas faster. It helps us innovate more collaboratively with our customers and provide better solutions to consumers at global scale. It is also helping us expand and solidify the partnership-oriented nature of how we go to market. Lastly, design thinking is helping us drive innovation throughout the broader commerce ecosystem where our scale has impactful change.

Figure 3: Design and creativity are becoming more critical to Mastercard's success in the digital world

As design thinking permeates the business world, there are evolving perspectives – not all positive. In the design community, perspectives are mixed as well. It is good to point out that there are clear pros and cons for organizations evaluating the applicability of design thinking and how to maximize outcomes in practice.

> "Design thinking helps us look beyond our own internal groupthink to get to new ideas faster. It helps us innovate more collaboratively with our customers and provide better solutions to consumers at global scale."

Pros include high engagement across product teams, breaking down silos, and overcoming muscle memory to see past execution challenges, bias, and inertia. Cons are that it can brand activities as human centric that are more 'inside-out' than 'outside-in', maintain mediocrity and require external expertise to land solutions in market. The relative maturity of the organization is key to optimizing the positive and minimizing the negative.

Design, or 'UX' more broadly in digital, is best as the third leg of the stool – with business and technology on the other two. Successfully seeding a design thinking mindset depends on the relative maturity of the other key parts of innovation development. If you come up with fantastic ideas that will take you years to execute, the expectations set via design thinking and commensurate high engagement can send you from peak to valley quite quickly.

To avoid this level of employee disillusionment, make sure that you have your technology teams on board. In our case, we seeded design thinking in our innovation lab, where our research and development engineers were able to take ideas from concept to execution rapidly. This embeds the possibility of driving material outcomes and allows time for broader scale technology modernization to take root. On the business side, executive sponsorship is key, as are measurable results and incentivization. Business leaders want to see that what can be a material investment will pay off. For us, again, our seeding of this in our innovation lab provided enough of the softer proof points to gain business buy-in.

Another key consideration for teams is the scaling of design. We are seeing a significant trend across leading technology companies of embedding designers in their organizations. For traditional software development, leading companies are working towards having one designer per ten engineers. In mobile development, that ratio is more concentrated with one designer per three engineers. By scaling highly skilled designers, efforts can be amplified with extensive knowledge of modern design patterns, deeply empathic practitioners who can lift-up broader team efforts, and talent that can better navigate the execution of challenges on the ground with the cross-functional team.

Mastercard is progressing here, as we are growing design capabilities across our product development teams. We are also scaling research capabilities to allow us to be more responsive to user feedback as we are prototyping and iterating our design solutions. In fact, three of our technology hubs have built-in consumer testing capabilities, while others use portable testing kits that can be set up in the field. We are embracing cloud-based insights platforms that extend our reach globally to field moderated and unmoderated studies. Together, these capabilities allow us to maintain high empathy for our users – the essence of design thinking – throughout our product development lifecycle.

WHY DESIGN THINKING MATTERS

ANNE-LAURE FAYARD | Associate Professor of Management, Department of Technology Management and Innovation, NYU Tandon School of Engineering

ABSTRACT

Design thinking has been an important management trend over the last decade and is still very much in fashion. Yet what design thinking really is and what it entails in practice for organizations often remains nebulous. In this article, I argue that design thinking is not a new concept and explain why it has aroused such a keen interest in recent years. I highlight the value of design thinking as an innovation process and stress the implications it has, as a mindset, for organizational culture. Lastly, I stress the need to use design holistically – with a system perspective – to develop meaningful and socially responsible innovations.

1. INTRODUCTION

In September 2015, Harvard Business Review released a special issue entitled "Design thinking comes of age." It covered the use of design thinking in multiple industries, reflecting a growing interest in the topic at the time, which has only multiplied as evidenced by the number of publications, case studies, and use of the phrase, as reflected by Google Trends. If the term is not new (early references to it can be traced back to the 1950s and 1960s), it has lately become part of the popular discourse and gained momentum in the business world as many businesses send their employees to training courses on design thinking and seek to incorporate it in their operations with the hope of nurturing a more innovative culture, boosting product and service innovation, as well as improving customer experience.

At the same time, some companies, such as IBM and SAP, have developed their branded version of the design thinking process, creating some confusion as to what design thinking is beyond the branded variations. In this paper, I review some of the origins of design thinking, discuss its main principles and methods, and illustrate it with examples. I briefly discuss why it has become such a favored approach (a question I am often asked) and argue that, to successfully implement it, organizational and cultural conditions – beyond Post-It notes and whiteboards – are required. Lastly, I emphasize the importance for design thinking to be holistic, systemic, and socially responsible.

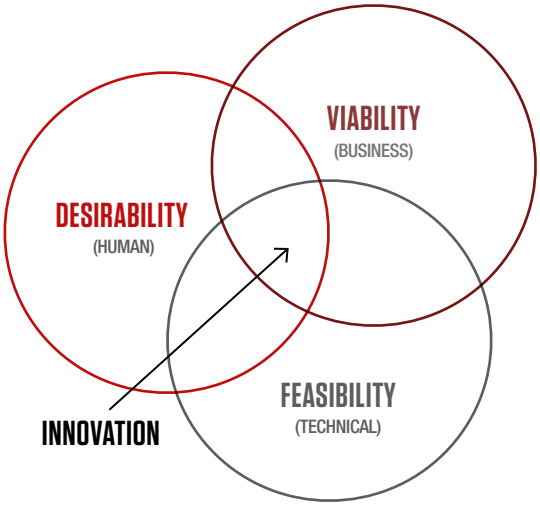
2. DESIGN THINKING: WHAT IS IT REALLY?

2.1 A brief history[1]

The history of design thinking goes back to the mid-1950s when Buckminster Fuller, a technologist, designer, and inventor, began teaching a course on design science at MIT. Fuller liked calling himself a comprehensive

[1] For more details on the origins and history of design thinking, please read: Dam, R., and T. Siang, 2018, "Design thinking: get a quick overview of the history," Interaction Design Foundation, https://bit.ly/2q02kYE; and Szczepanska, J., 2017, "Design thinking origin story plus some of the people who made it all happen," Medium, January 4, https://bit.ly/2taChUM

Figure 1: Design thinking: a human-centered approach to innovation



> "Design thinking is a human-centered approach to innovation that draws from the designer's toolkit to integrate the needs of people, the possibilities of technology, and the requirements for business success." — TIM BROWN[8]

anticipatory design scientist – "an emerging synthesis of artist, inventor, mechanic, objective economist and evolutionary strategist."[2] Herbert Simon, with his seminal text, The sciences of the artificial (1969),[3] also attempted to develop a science of design and consequently left out concepts like intuition, experience, and social interaction from his definition of design.

Around the same time, participatory design, an approach that also deeply influenced design thinking as we know it today, emerged in Scandinavia. More specifically, participatory design, which focused on social interactions and organizational contexts, influenced many of the advancements in human-computer interaction and service design, as well as the focus on user-centered approaches and co-creation. At the core of the Scandinavian participatory approach was a desire to improve our understanding of people's work to create better information systems, a willingness to involve users in an effort to reduce resistance to change, while increasing workplace democracy.[4]

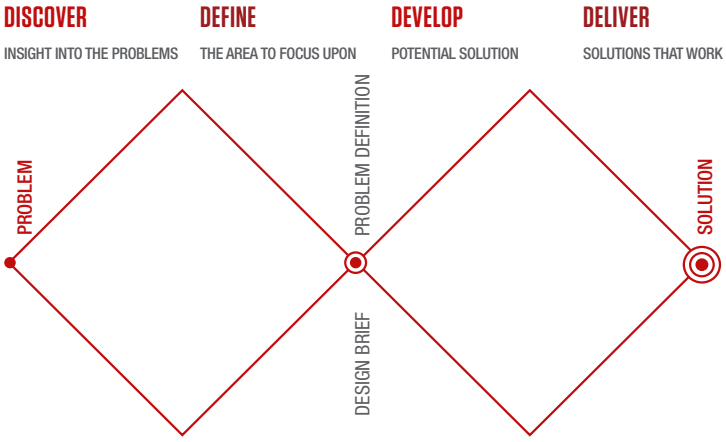
This approach has been slowly adapted and embraced (though its influence has often been forgotten) by designers in North America and other European countries. Researchers like Lucy Suchman, an anthropologist who joined Xerox's Palo Alto Research Center (PARC) in the late seventies, brought back the ideas of codesign and user-centered design to North America. In the 1980s, Nigel Cross[5] and Donald Schön[6] studied designers and their practices and highlighted specific dimensions of designers' ways of solving a given process, such as the rapid generation of multiple solutions, a bias towards action rather than prolonged analysis, and the iterative nature of the process. In these studies, design emerged as a practice closer to arts and crafts than to science.

While design thinking is rooted in the work practices and tools used by professional designers, it is a way of thinking not limited to professional designers or architects. On the contrary, it proposes an expansive definition in which everyone is a designer, and anyone can design. "Design thinking" refers to a view of design as an approach to problem solving that is deeply exploratory and where the problem is not considered a given, but one that must be defined and redefined through exploration. That approach is what Boland and Collopy (2004)[7] call a design attitude, which they contrast with a decision attitude. A decision attitude, taught in most schools – particularly business and engineering schools – presupposes that the problem provided is well defined and that a number of alternative solutions exist for it. The remit for people trying to tackle problems is then to figure out the best solution among a set of alternatives often represented by complex modeling systems. Yet such a decision attitude seems to have reached its limits in light of the complexity of problems faced by firms and society. In turn, companies seem to find value in designers' approach to problems.

Hence, recent years have energized managers, among other professionals, to engage in design thinking, described as a more human-centered approach whereby designers "match people's needs with what is technologically feasible and what a viable business strategy can convert into customer value and market opportunity" (Tim Brown[9]).

[2] Buckminster Fuller Institute, https://bit.ly/2xWJ7NO
[3] Simon, H., 1969, The sciences of the artificial, MIT Press
[5] Cross, N., 1982, "Designerly ways of knowing," Design Studies 3:4, 221–227
[6] Schön, D. A., 1983, The reflective practitioner: how professionals think in action, Basic Books
[7] Boland, R. J., and F. Collopy, 2004, "Design matters for management," in Boland, R. J., and F. Collopy (eds.), Managing as designing, Stanford Business Books
[8] Brown, T., "Design Thinking), IDEO, https://bit.ly/2Ogn5zB
[9] Brown, T., 2008, "Design thinking," Harvard Business Review, 86:6, 84-92

Figure 2: The design double diamond



Source: UK Design Council, https://bit.ly/2zJzFwS

Design thinking (also known as human- or user-centered design) is usually associated with creative organizations like design firms. Originally, design thinking was indeed the domain of design and innovation consultancies, such as IDEO and Frog Design, that companies would hire to help them develop innovative products and services. Interestingly, in the last decade or so, companies in multiple domains have been bringing design thinking in-house and developing internal capabilities. For example, healthcare has welcomed multiple projects that aim to create better experiences for patients. Mayo Clinic started a pilot project in 2002 where patients, physicians, and designers collaborated to generate ideas on how to improve the doctor-patient interaction. In 2007, this pilot was integrated into the clinic's new Center for Innovation.[10] In the field of technology, multiple companies – Samsung, Intel, SAP, and IBM, to name a few – have included design-thinking methods in their innovative processes. In the banking sector, BBVA and Deutsche Bank are sending their employees to design thinking training and developing internal capabilities with innovation labs, and Capital One acquired design consultancy Adaptive Path in 2014 and is trying to develop design thinking skills internally. Additionally, in the last five years, management consultancies have begun acquiring design consultancies; for example, in 2013, Deloitte acquired Doblin, Accenture acquired Fjord, and in 2015 McKinsey acquired Lunar. More recently, these management consultancies began building their internal capabilities and staffing their own design team.

[10] Kersten, P., "Inspiring innovation throughout the healthcare industry" Center for Innovation, https://mayocl.in/2NQa5BC

[11] Case study: Bank of America, Bloomberg Businessweek, June 19, https://bloom.bg/2zDnu79

2.2 The design thinking process

The process comes in many different forms — three, five, or seven steps depending on the school of thought and company. I personally find the "Double Diamond" developed by the U.K. Design Council a useful way to visualize the process.

Although it suggests phases, the double diamond does not clearly delineate the steps practiced, suggesting a more fluid and dynamic way to engage with the process. It also highlights the divergent and convergent modes involved in the process: during the divergent mode, you take an exploratory stance, research, and brainstorm widely; and during the convergent mode, you make sense of results (research insights, ideas, prototypes, etc.) from the divergent phase, narrow your options to a few, and then choose one that you can implement.

Underlying these phases are a few key principles and assumptions regarding where new ideas come from and how to evaluate them:

1. Reality is complex and socially constructed. Consequently, it is important to **gain deep empathy for users' or customers'** needs and aspirations for the future. This means not only understanding their current needs (especially because they might not always be able to articulate them) but also **the broader context that shapes their daily lives.**

When designers at consultancy IDEO started research for a project that sought to inspire people to open new bank accounts with Bank of America, they began by gaining empathy for different types of users. They noticed that many individuals often rounded up their financial transactions. It made the math easier and created a "buffer" (of extra money) on their bank account. The IDEO team also noticed that many of the users they observed and interviewed had difficulty saving what money they had, whether due to a lack of resources or willpower. This led them to come up with Keep the Change, a new service that rounds purchases made with a Bank of America debit card up to the nearest dollar, with the difference transferred from checking to saving accounts. It launched in 2005, and in less than a year, it attracted 2.5 million customers, leading to the opening of more than 700,000 new checking accounts and 1 million new savings accounts for Bank of America.[11]

2. Ideas can only be assessed **in practice**. Therefore, a **bias towards action** and **prototyping** are central

to design thinking, a misleading term because design thinking is as much making as it is thinking. You need to test your assumptions by generating multiple prototypes and getting feedback from users. Prototyping creates opportunities for learning, which means being ready to let go of your "great idea."

This is what happened to the Air New Zealand team that worked with IDEO to redesign its long-haul flight experience. Reeves, Air New Zealand's program director, explains how the team had come up with a new seating idea: "We thought it was a perfect idea and were extremely excited about it. We built a cabin of seats and tested them and all the passengers hated it... It was such a powerful outcome for us and something we wouldn't have learnt, had we not prototyped."[12] And prototyping helped the Air New Zealand team avoid spending large amounts of money on the redesign of new seats.

You can prototype anything, including a service, experience, or program, allowing you to learn as much as with physical prototypes. For example, you can rent a space and furniture and invite users to experience your new service. In my research on service designers, I compiled multiple examples of service prototypes in banks, airports, and hospitals. Prototyping is also very powerful for social innovation as described in the example below.

With a group of students,[13] we worked on a social innovation project, the Bindi Project,[14] a community-centered program that aimed to empower women from underserved urban areas. To test our original idea, we collaborated with a Nepalese NGO and piloted our program with 36 women in a slum of Kathmandu.[15] During the pilot, we prototyped multiple components of the program. For example, to test our assumptions that women would be willing to give back with their time to train other women and share what they learned during the program, we ran a workshop in a slum, where we trained women on sexual health and then asked them to volunteer to run a similar workshop in their community. We ended up with several participants signing up to lead future workshops, suggesting that our assumption about the willingness of participants to give back to the program was (at least partly) substantiated.

3. Ideas never arrive fully formed. To develop a service or product that truly fulfills the needs of your users or customers, you need to **iterate**, and prototypes allow you to refine your ideas and flesh them out.

Intesa Sanpaolo, a major Italian bank, worked with Frog Design to design a new branch for their customers with the goal of transforming the relationship between the bank and its clients beyond being purely transactional. "To achieve this, over 600 clients and 300 employees engaged in test interactions in a full-scale branch prototype, as the design was refined in real time to better serve their needs."[16]

R & D Lab Sproutel similarly iterated on prototypes to create Jerry the Bear,[17] a teddy bear with type 1 diabetes. Jerry helps children with the same diagnosis learn more about their condition and feel less alone. It took the Sproutel team 29 iterations and multiple workshops with children to develop the bear, which began retailing in 2017.

4. Acknowledge that **innovation is collaborative** and requires different skills and expertise, which is why **multidisciplinary teams** are crucial. In diverse teams, problems are seen from multiple angles, and new solutions arise from the merging of these different perspectives. It is about letting go of the myth of the lone inventor. Creative collaboration also means **co-creating with your customers and other stakeholders**. They are experts in their practices and must be involved in the process – at the very least in the prototyping and testing phase. Engaging them earlier in the process can also be generative.

In participatory design projects, such as one I conducted with air traffic controllers,[18] users have a deep expertise, and you cannot develop tools without learning from them and engaging in the process. However, children suffering from diabetes are also experts in their needs and experience, and, because the Sproutel team acknowledged their expertise and engaged them in the design process by inviting them to multiple prototyping workshops, Sproutel was able to successfully develope Jerry the Bear.

5. Recognize that some innovative solutions may not be flashy or super complex. In fact, good design is often

[12] Lakhani, K. R., A.-L. Fayard, N. Levina, and S. H. Pokrywa, 2012, OpenIDEO, Harvard Business School Case Study 9-612-066, p. 5
[13] This was a project with Design for America of NYU, a student organization that used human-centered design to tackle social issues, for which I am a faculty advisor.
[14] The Bindi Guide, https://bit.ly/2OUL1FP
[15] The Bindi project, https://bit.ly/2R5lj1G
[16] https://bit.ly/2N7ApS9
[17] https://bit.ly/2xVz0bW
[18] Mackay, W. E., Fayard, A-L, Frobert, L. and Médini, L. 1998. Reinventing the Familiar: Exploring an Augmented Reality Design Space for Air Traffic Control. CHI 1998: 558-565

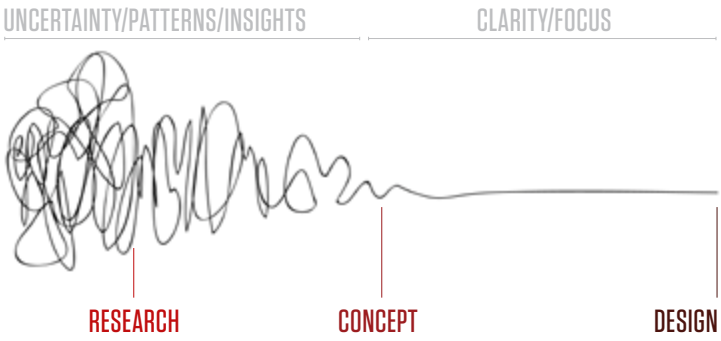
simple and "invisible" to users, allowing them to do their work or continue their daily activities in a seamless, improved manner.

If you think of Air New Zealand, their passengers wanted comfortable seats that did not require spending ten minutes reading an instruction manual or even watching a three-minute video. In 2007, Engine Service Design, a London-based service design and innovation consultancy, worked with Virgin Atlantic to help redesign its passenger experience at Heathrow Airport.[19] *During the project, Engine designers noticed that, for many people, the first ten meters into the airport are the worst. To reduce passengers' stress, Virgin Atlantic positioned staff members near the terminal's entrance to greet people and provide them information when needed. This is an example of a simple innovation sparked by observing users and one that improved passengers' experience, as well as reduced confusion within the terminal.*

3. THIS IS GREAT! HOW CAN WE EMBRACE DESIGN THINKING IN MY COMPANY?

The potential of design thinking to generate innovations that can bring value to users and organizations has been demonstrated by many case studies similar to the stories I shared above. Yet, being convinced of the potential of design thinking is not enough to implement it effectively, and people often ask me how they can apply design thinking in their work and/or implement it in their organizations: What tools do they need? What kind of training is recommended?

Figure 2: Damien Newman's design process squiggle



Source: That squiggle of the design process, ReVision Lab, https://bit.ly/2KMyRMM

They often worry that they or their organization is not creative enough. Indeed, design thinking is often associated with cool and creative spaces with whiteboards and tons of sticky notes where teams (multidisciplinary if possible) brainstorm new ideas. However, if sticky notes, Sharpies, and whiteboards are useful, they are not the sine qua non condition of design thinking. Furthermore, design thinking can be accomplished without sticky notes and whiteboards. Similarly, if you think that using one of the phased processes of design thinking is the silver bullet for new ideas and innovative solutions, you will again be disappointed.

Design thinking is not a science nor a magic recipe. It is an art and craft that requires a certain mindset or ethos.[20] This mindset is deeply connected to the principles of design thinking listed in the previous section: it is about being empathetic, holistic, collaborative, and willing to experiment and iterate. Intrinsic to a design thinking mindset is also the ability to embrace ambiguity and to accept that the process may be messy, and the double diamond might look more like a squiggle (see figure 2) than a nicely delineated double diamond or phased process.

Mechanically following the steps provided by various design thinking frameworks or putting Post-It notes on walls and developing journey maps (or any other tools used in design thinking) will not guarantee lightning bolts of innovation or interesting new ideas. Indeed, for the design thinking mindset to be fully embraced by organizational members, it is crucial to have a culture that supports it. More specifically, you need an organization where collaboration, experimentation, risk taking, and a user-centric approach are not just values posted on the walls or website. Moreover, you need to ensure that your innovation teams can get their ideas implemented. Generating ideas is not always the main problem; implementing them is far more difficult. The obstacles to implementation often include the organization's inability to execute in an agile, iterative way; anemic leadership support for funding "concepts;" and the inability to convince various stakeholders. This is what Samsung's leadership realized: to take advantage of design's full value, they needed to make sure that the design team would not come up with new products that would be contested or, at best, ignored by engineers and marketers. This is why the company embarked on a radical transformation of its culture. Samsung understood that if it did not, none of the new products would be produced.[21]

[19] Fayard, A-L, I. Stigliani, and E. Williams, 2010, "Designing services at engine," case study, Imperial College Business School

[20] Fayard, A-L., I. Stigliani, and B. Bechky, 2017, "How nascent occupations construct a mandate: the case of service designers' ethos," Administrative Science Quarterly 62:2, 270-303

Because design thinking is an approach for generating new ideas, it is important to make sure that your organizational innovation culture (e.g., where do new ideas come from and how to evaluate them) is aligned with the design thinking mindset.[22] For example, do you believe that empathy toward your customers or users will inspire the development of new products and services? Or do you think that using complex modeling and large datasets is more effective? While conducting ethnographic research at Nokia in 2009, Tricia Chang discovered an insight that challenged Nokia's business model of developing expensive smartphones for elite users and cheap smartphones for low-income users. She found that low-income consumers were ready to pay for more expensive phones, a finding that suggested revisiting Nokia's business model. However, Nokia management did not know what to do with her insights, which seemed weak to them because Chang had "only" 100 interviews in comparison with their sample size of several million data points. Chang argued that it was not surprising they could not see any of her insights "show up in their quantitative datasets because their notion of demand was a fixed quantitative model that didn't map to how demand worked as a cultural model in China."[23] Nokia ended up ignoring her findings. This is a cautionary tale for businesses on how relying too much on numbers, and ignoring data that was not easily measurable, may have contributed to Nokia's decline.

While Post-Its, Sharpies, and whiteboards are useful tools, they do not ensure the successful implementation of design thinking in an organization. Understanding an organization's culture and being ready to change it to support a design thinking approach is crucial.

4. DESIGN THINKING WITH A GRAIN OF SALT: IT NEEDS TO BE HOLISTIC, SYSTEMIC, AND SOCIALLY RESPONSIBLE

As innovation has become a strategic imperative for companies, taking a design thinking approach seems like a perfect model for developing new products and services that will increase customers' satisfaction and loyalty. Yet these innovations tend to be quite self-centered and local. They ignore the not-always-positive impact the innovations may have on other humans, other systems, and common resources like water, food, and our climate.[24]

While design thinking provides the tools for creating innovative, meaningful solutions with social impact, it is important to remember the unintended consequences that can result from the products, technologies, and services we design. These can emerge even when a design thinking approach has been fully embraced. Indeed, as you focus on end users or specific customers and create products and services that will provide great solutions for them, you may create problems for other groups of users locally and/or globally, other groups whose voices have not been taken into account. New products and services also have physical consequences on our environment (locally and globally) that we may not realize until after the fact. Consequently, asking about an innovation's impact from a system perspective is imperative. For example, car-sharing apps like Uber or Lyft, which are meant to simplify our lives (and often do in many ways), are disrupting urban infrastructures by increasing traffic congestion and reducing the use of public transportation. They are also disrupting employment, not just creating new and flexible job opportunities. Realizing the negative effects of their addictive design, former designers at Google, Twitter, and Facebook have disconnected themselves from the Internet.[25] I would argue that, in this case, user-centered design focused only on increasing usage rather than understanding contexts and meaningful interactions. Moreover, these designers did not adopt a system view that would facilitate their understanding of the social impact of the technologies they created and the practices those technologies generated.

While predicting all the unintended consequences of design is impossible because systems are deeply interconnected at a global level, it does not prevent designers, as well as companies that are increasingly using design thinking, to be mindful of these consequences and to be ready to challenge a new product or a service if it has negative effects in another part of the system. Hence, embracing design thinking should incorporate a system view and be socially responsible. This is not just an individual responsibility, but also one for companies, governments, and national and international institutions.

[21] Yoo, Y., and K. Kim, 2015, "How Samsung became a design powerhouse," Harvard Business Review, September
[22] Fayard, A-L., E. Gkeredakis, and N. Levina, Information System Research
[23] Wang, T., 2016, "Why big data needs thick data," Medium, January 20, https://bit.ly/23E9qlv
[24] Frick, P., and C. Luebkeman, 2017, "Planet-centred design: a mindset shift for engaging complexity," Huffington Post, January 19, https://bit.ly/2OWQsnA
[25] Lewis, P., 2017, "'Our minds can be hijacked': the tech insiders who fear a smartphone dystopia" The Guardian, October 6, https://bit.ly/2yMqOcH

THE ADOPTION AND IMPACT OF DESIGN THINKING IN FINANCIAL SERVICES

PAUL LEE-SIMION | CEO, AA INFO, and Senior Consultant, DBS Singapore

ABSTRACT

The adoption of design thinking techniques is not an option for financial services, it is a must. The landscape is changing rapidly, and design thinking is affecting everything we do and use; the business case for it is indubitable. Governments are paying attention and adopting the approach, as well as showcasing the pronounced changes fintech is having on the financial industry. By adopting design thinking, the financial industry is undergoing a paradigm shift in the way they engage, organize, and deliver to their customers.

1. WHAT IT IS DESIGN THINKING?

Design thinking is affecting everything we do and use. Design thinking, as we know it today, was developed from the creative thinking of the 1950s, when mechanical engineers would add something to their products to make them distinctive with the intent of making them "less threatening" to customers. An iconic example of this is the coke bottle (Figure 1). Today, design thinking connects users with products for a meaningful experience.

Terms such as agile, customer centric, customer journeys, re-imagining, innovation, and infographics are all associated with design thinking. Design thinking is about delivering meaningful solutions to problems, that is change with impact.

The change can be a process, a product, an environment, or an artefact. The delivery itself may also change. Ensuring that those affected by the change have an affinity to the solution is fundamental to design thinking. Achieving this is more than about understanding the users and the impact. It involves engaging those impacted right at the beginning of the process, and iterating outcomes. Apple's iPhone is a great example. The iPhone facilitated the change in how music is delivered, how photographs are taken, and how the world communicates. This product is also updated every few years because clients want a refresh.

It is because of design thinking that Apple went from a near defunct company to what it is today. Governments and other companies have noticed and are engaging with the concept as well. Apple has a long history with the design group Frog in aesthetics and usability of its products. Frog have also assisted UBS in the design of UBS's mobile app to "make its clients' personal financial information easily accessible, easily readable, and personally meaningful."[1]

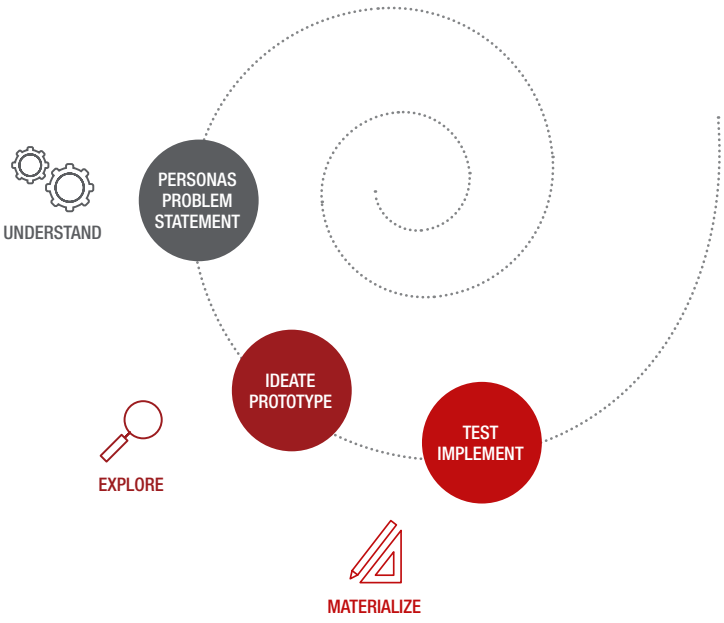
Upfront, design thinking focuses on the sensitivities of those who are the subject of the challenge one is resolving.

[1] https://bit.ly/2Ofk3fM

The issue being resolved is the "problem statement" such as, "how can every vehicle become a taxi." Those initially perceived to benefit from the outcome are the personas. Personas are imaginary people who develop, change, and grow with the ideas portrayed, such as the drivers and passengers. Going through several iterations, one ends up with several "opportunities" to resolve the task in hand and a reasonable idea of the group of people who will benefit from it. This process is always enlightening because anything is initially in bounds, the most outrageous ideas welcome and the perception of how your personas will react is surprising.

Figure 1: The changing faces of the Coca-Cola bottles

Figure 2: The design thinking spiral



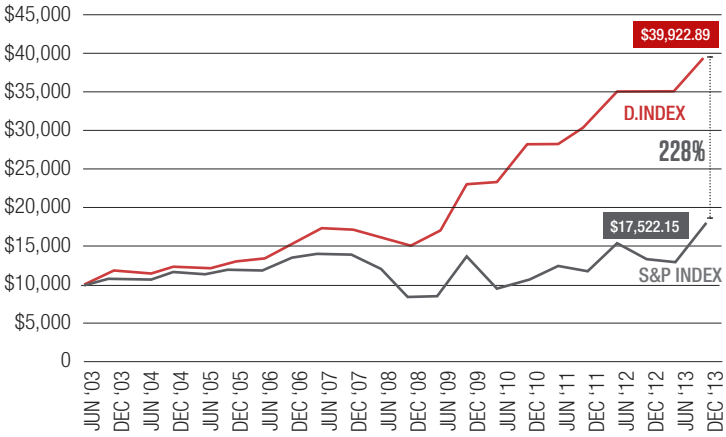
Finding the most "impact problem to solve," is very relevant to today's enterprises. The design thinking approach starts with prioritizing issues to attack and focuses on the ones that can help deliver the most value to the organization. Solutions are then proposed, empathized through the personas, and quickly iterated.

The practicalities come next. Some opportunities simply cannot be delivered. Timescales, technology, and budget dictate that these need to be put aside for later. At one design thinking workshop, the opportunity proposed was a dongle that would project a hologram so that clients could communicate with their bankers. The problem statement was "how can bankers spend more time with their clients." The practicable opportunity delivered was artificial intelligence that "listened" to conversations and produced an actionable list of tasks for the banker. (A further development was to automatically execute the actions on the list where possible.) This impact allowed the bankers to increase the amount of time they spend with their clients by 30%. The hologram opportunity was put aside for later.

Representative users are exposed to the opportunities emanating from the design thinking process; this is akin to bringing your persona to life. The insights from the users could even rephrase the problem statement and then iterate again. Opportunities are prototyped. Prototypes come in many forms, users experience them, and the personas adapt with the feedback. More ideas are generated, and opportunities develop. We spiral round and round with the opportunities that have been encapsulated by the prototype flying out (Figure 2).

Only then is production or implementation considered, and at all times the impact on the persona assessed. Leaving the business benefits of design thinking to one side, this way of resolving issues is fun. It ticks all the boxes, including all stakeholders, especially those often forgotten, who tend to those most impacted by the change.

Figure 3: DMI – design value index



Source: Westcott and Rae (2014)

DESIGN-CENTRIC ORGANIZATIONS:

APPLE
COCA-COLA
FORD
HERMAN-MILLER
IBM
INTUIT
NEWELL-RUBBERMAID
NIKE
PROCTER & GAMBLE
STARBUCKS
STARWOOD
STEELCASE
TARGET
WALT DISNEY
WHIRLPOOL

2. ECONOMIC IMPACT OF DESIGN THINKING

Design thinking has been applied with tremendous impact in many industries, including consumer products, education, financial services, healthcare, NGOs, and transport. Examples include:

- **Consumer products:** over 2 billion iPhones, enabling Apple's profits to increase to over U.S.$50 bln annually from U.S.$3.5 bln.
- **Education:** classroom layouts designed with the needs and desires of pupils taken into consideration, resulting in pupils being far more engaged. Fidget modules increase attention span of eight year olds from 7 to 20 minutes.
- **Financial services:** moving banks to realize that their customers dictate the way business is to be done. 150% increase in new investment product sales.
- **Healthcare:** moving beyond controlling pain and taking care of medical emergencies – U.S.$400 bln market in U.S. alone.
- **NGOs:** 500% increase in meals delivered.
- **Transport:** increased passenger throughput by repositioning and synchronizing trains, escalators, barriers, and ticket offices accommodating 5% year on year growth in passenger capacity.

Analysis by Westcott and Rae (2014)[2] clearly shows that design-centric organizations benefited from a 228% share price uplift over a 10-year period. Further analysis indicates that if it was easy, everyone would be doing it – only 1/5th of U.S. listed companies have actually benefited from the uplift, as compared to those who proclaim to practice design thinking.

The question every CEO, shareholder, and government asks is: "how do we make money?"

To make money, the institution first and foremost needs people to buy their products (or use their services). Second, the institution needs to be maintained by people continuing to buy its products. Design thinking changes the paradigm from products and services being purely functional to a user experience that people perceive they need.

Henry Ford paraphrased the cycle:

1909 – about the model T: "You can have any color as long as it's black"

1927 – "Failure is only the opportunity more intelligently to begin again"

Design thinking brings failure upfront. The product solves a real problem and in so doing people will want the product. Then, the institution moves onto the next real problem. Consequently, upfront, design thinking identifies and validates the problem, positions that with potential customers, and puts opportunities on the shelf that the customer is unlikely to buy into or cannot be delivered at that time.

Developing products to which customers will have an affinity saved GE U.S.$15 mln in development costs when they first took this approach in 2010.[3] It also allows maintenance costs to be halved.[4]

[2] Westcott, M., and J. Rae, 2014, "Design-drive companies outperform S&P by 228% over 10 years – the DMI design value index," Design Management Institute, March 10, https://bit.ly/2LTnT9G
[3] Forrester Customer Experience Marketing
[4] Silver AG/Lucerne University of Applied Science and Arts 2017

Design thinking also allows for value-based pricing. The manufacturing costs of an iPhone X is U.S.$370 and the customers will pay U.S.$1000 [Richter (2017)[5]]. This has contributed to Apple becoming a trillion dollar company. Apple spends U.S.$10 bln annually on research and product development, which is only 0.1% of Apple's market capitalization. Design thinking is at the heart of everything Apple does and is commercially effective. Who would not spend 0.1% of their market capitalization to be the world's first trillion-dollar company?

3. GOVERNMENTS AND DESIGN THINKING

While companies that embrace design thinking have been found to outperform those that do not, the manner in which governments embrace and support it creates an innovative culture in their country to attract more companies (and, therefore, add to their revenues), as well as promoting social cohesion, decreases in taxes, and saving lives by delivering services with affinity to their citizens, who after all pay for it.

Design thinking by the Singapore government has had immense impact, including:

- **Law and order:** making courts family friendly with children's play areas and upbeat décor instead of a somber environment.
- **Healthcare:** all disciplines for patient groups (e.g., the elderly) around the same "island," that is on the same floor with practitioners spreading out from the center. Results are a 40% increase in daily patient treatment and freeing up space.
- **Town planning:** minimizing the distance anyone has to travel for public transport, health, social, and shopping facilities, including making the journey accessible to the disabled.

In Canada, the Canada Beyond 150 program has introduced "cultural probes." These engage people through their own recordings, drawings, notes, and artefacts to make them comfortable in portraying their insights to the government. Through their Employment and Social Development Innovation Lab, Canada has bought together people from various disciplines to simplify the application to and running of the Canada Pension Plan.

MindLab in Denmark was the first government innovation lab, established in 2002. However, Mindlab will close down at the end of this year with activities going to Denmark's Disruption Task Force. Notable outcomes from MindLab include driving the Danish government's digitization program and knowledge transfer to the U.S., U.K., and Mexican governments in setting up innovation labs.

U.S.: there is probably no other organization in the world other than the U.S. government that has the power to improve lives. Yet while design thinking is prominent in the private sector and intergovernmental agencies, it is less prominent in U.S. agencies. Some federal agencies have had success in the specific context in which design thinking has been applied. These include:

- Call centers eradicating the need for the public to know every form and detail in order to get information by having the customer state what they want and the burden passed to call center staff equipped with natural language search technology.
- Re-imagining the U.S. Air Force, under the slogan "doing smarter stuff faster."
- Simplifying federal tax forms.

The South Korean Government has taken the approach of immediately partnering with commercial organizations to leverage the technology available. Outcomes with SAP include:

- An Internet of Things (IoT) solution to monitor livestock feeding.
- A public service design model, increasing participation and literacy rates.

The U.K. Government Digital Services are delivering services through .gov.uk. This includes "verifying," which uses third-party data to determine who you say you are in order to access services through .gov.uk

Brazil has had its recent challenges. Prior to these challenges and throughout them, all administrations remained committed to disclosing detailed information about federal spending – and giving citizens a clear view of their tax dollars in action.

Given that governments are beginning to understand the impact of design thinking, it is not surprising that they are paying particular attention to the areas in which it is applied, one being fintech. Government support,

[5] Richter, F., 2016, "The cost of making the iPhone X," Statista, November 16, https://bit.ly/2A3NkI0

especially in Australia, Canada, Hong Kong, Singapore, South Korea, China, the U.K., and U.S., is through regulators and industry associations. Singapore's Monetary Authority hosts the world's largest fintech festival in November with over 500 start-ups from over 40 countries.

> "Institutions that are adopting design thinking will retain their clients and gather more. Those that do not will become extinct."

4. ADOPTING DESIGN THINKING IN FINANCIAL SERVICES

Companies such as PayPal, WoldRemit, WeChat, FWD, and Uber continue to challenge traditional thinking in retail financial and insurance services. In the wholesale financial market, companies such as Rippel may give Swift a run for their money. iRX is changing the wholesale insurance paradigm and companies such as Aquan/Zilliqa are providing platforms, reaching U.S.$1 bln market capitalizations.

The history of these fintech companies is one of being outrageously disruptive and, in particular, embracing design thinking. Traditional financial services organizations are, therefore, endorsing the approach in order to be part of their clients' life as opposed to an unloved necessity. Examples include:

- One Singapore bank re-imagined their process of explaining products to clients, resulting in increased sales of its new investment product by 150%.
- A CIO of a global European bank defines innovation as a change in process and deploys design thinking to reduce costs by over 30%.
- An Australian insurance company increased profits by 35% by enabling interaction with clients to complete necessary tasks within 60 seconds.
- At a recent conference, Citibank suggested that it takes three years to build a mobile only bank. DBS took nine months to build one in India, which has over 1.2 million customers [Marous (2018)[6]]. Standard Chartered have announced they will build a digital bank in Hong Kong.

The problem statement for traditional financial services is around how to embed themselves in the client's daily life. To do so, they need to make delivering financial services not only fun but also seamless and subliminal, and embed it in customer journeys.

Consequently, financial services firms need a paradigm shift in understanding their clients, rapidly testing (and failing) products, and bringing the products to market within months. Everything the financial institution does now needs to have the client at the center and need to be continually evaluated by clients and personas. Success is having the customer availing themselves with the products and services provided without necessarily knowing that they are doing so. The product finds the clients when and where the client needs it. Within financial institutions, this requires a radical transformation in processes, controls, regulation, protection, technology, and delivery.

The manner in which financial institutions organize to deliver (and fail) rapidly varies. In all cases, financial institutions adopt a design thinking approach:

- Customers are continually surveyed by their current interactions with the financial intuitions (whether branches, agents visiting, online, or mobile), as well as directly through discussions. Their aspirations and challenges need to be understood and problem statements need to be formulated and personas developed.
- Teams involving people from all disciplines including user experience-experts, process, and support are assembled to come up with several ideas and to understand how these will affect the customer through developing several customer journeys.
- A prototype of the opportunities that will have the most impact is then developed in days or weeks and the user-experience is assessed. One or more opportunities are then agreed upon to be delivered as a product.
- The team is expanded to deliver continuously (at least daily), releasing features, and assessing the released features against the persona, clients, and those affected. Where the features are not well received, the user-experience is captured, adapted with the persona, and the change delivered immediately.

[6] Marous, J., 2018, "How long should it take to build a digital bank?" The Financial Brand, https://bit.ly/2FCgDM1

5. THE IMPACT OF DESIGN THINKING ON FINANCIAL SERVICES

Institutions that are adopting design thinking will retain their clients and gather more. Those that do not will become extinct.

Banks are continually redesigning their branches. Now they are doing so for the customer experience instead of for throughput and staff efficiency/metrics. This has even developed into different brands, such as Frank by OCBC in Singapore for certain demographics. Other demographics want the social experience of being in the branch with the banker coming to them and delivering their transactions; for example, a cash withdrawal over a cup of coffee. Each branch can be different to cater for the local demographic they serve. The revamping of the branch takes weeks instead months.

The use of smart phones is where the impact of design thinking has been most pronounced for financial services. Not only are the traditional providers being challenged, they have technical debt that inhibits them to move quickly. Furthermore, traditional providers think like and have regulated processes of traditional providers. Their transformation has to be radical in technology, manner of organization, use of data, and delivery.

The adoption of design thinking techniques is not an option for financial services, it is a must. The landscape is changing rapidly and they need to adapt.

THE DESIGN THINKING FALLACY – ARE BANKS IMMUNE TO INNOVATION?

ARJUN MURALIDHARAN | Principal Consultant, Capco
NIKOLA ZIC | Consultant, Capco

ABSTRACT

This paper examines why the financial services industry may be considered a laggard in innovation compared to other industries such as commerce or transportation. Inherent factors such as risk culture may negatively impact the overall ability to innovate. In this paper, we look at the potential of 'design thinking' to help financial institutions become more innovative, proposing a way forward to embed this innovation methodology effectively within a financial services company.

1. THE STATE OF INNOVATION IN FINANCIAL SERVICES

1.1 Concepts of service innovation in financial services

Leaders of most service businesses find little guidance in existing studies of innovation. The central themes of R&D, intellectual property, and breakthrough technologies often miss how service businesses evolve by steadily generating and implementing new ideas. The lack of guidance would not be puzzling if services' share in the business sector were small, or if innovation in services were unimportant. However, neither is true. In modern economies, service businesses account for most of the value created. In the U.S., for example, services now account for about 78% of GDP; the major economies of Europe and Asia are not far behind.

Even the manufacturing sector, which accounts for most of the remainder, incorporates significant services in the products it creates.

Innovation in services is important in part because it is one of the only effective ways to fight commoditization. Forces behind commoditization of services are fierce, and getting more so, as these markets become more open, more tradable, and more contested. Commoditization often occurs even faster in services than in physical products because innovations are easier to copy, there are fewer patent protections, lower front-end capital investments, and shorter product cycles. The rapid rise in global services trade, the significant liberalization in cross-border flows of services and capital, and the rapid globalization of many service firms are evidence of this trend. In many industries, for example, compensation for providing intermediation – the services of middlemen – has collapsed.

Despite these distinct trends, writing on innovation remains primarily focused on physical products and high technology.

We define innovation broadly as the combination of creativity and implementation. Thus, we focus on both the production of novel and useful ideas that improve effectiveness, as well as the methods used to put the creative ideas into practice. Innovation can include doing old things in new ways rather than developing completely

new inventions, which includes ideas originating from outside the organization that are customized to an organizational context or clientele. New ideas must be implemented – i.e., delivered to customers – to create commercial value.

Competing on service innovation requires a more intensive set of organizational practices than competing on physical product innovation. To understand these differences, we examine the practice of design thinking as a method for service innovation, applied to financial services.

There are service innovation pitfalls that arise from the unconsidered application of concepts conceived in a physical product innovation context. For example, research on physical product innovation tends to focus on radical, game-changing shifts, whereas innovation in services tends to be more fluid and evolutionary, and thus top competitors are characterized more by their steady pace over time than by making gains with long-shots. New product innovation in manufacturing involves significantly more fixed investment and greater commitment to longer production runs, making it necessary to move things more abruptly to justify these investments. Similarly, in new technology innovation, what is often at stake are new industry-wide standards and infrastructure, which tend to be more discrete. Exploring the differences between product and service innovation illuminates the importance of fostering a 'service innovation culture,' which we define as the consistent, coherent, and comprehensive presence of values and norms that promote fresh thinking and swift execution in service firms. Organizational structures and processes are the building blocks of this culture, and they include formal and informal incentives, socialization, role modeling, and venues for sharing information. The behavioral norms and values that define culture are vitally important in services, in contrast to physical products, because behavior itself is the product.

1.2 Design thinking – a five-minute primer

Design thinking is a service and product innovation methodology based on an iterative process between research, development, testing, and reengineering with a constant focus on the final user's acceptance. The methodology dates back to the 1960s, where the ideation approach was created in Palo Alto at the Stanford University. Today it is being applied in a global network of universities and workshops and practiced in all kinds of industries. Physical products such as the iPhone are likewise design thinking offspring, as are services such as AirBnB and Uber. Design thinking means many things to many people, and this pluralism persists into the practical implementation as well. There are a wide variety of process breakdowns and visualizations ranging typically between three and seven steps. Each phase, such as observing or testing, embodies one or more of the core ingredients of design thinking, such as empathy, reframing, ideation, prototyping, and testing.

FOUR THINGS DESIGN THINKING IS NOT

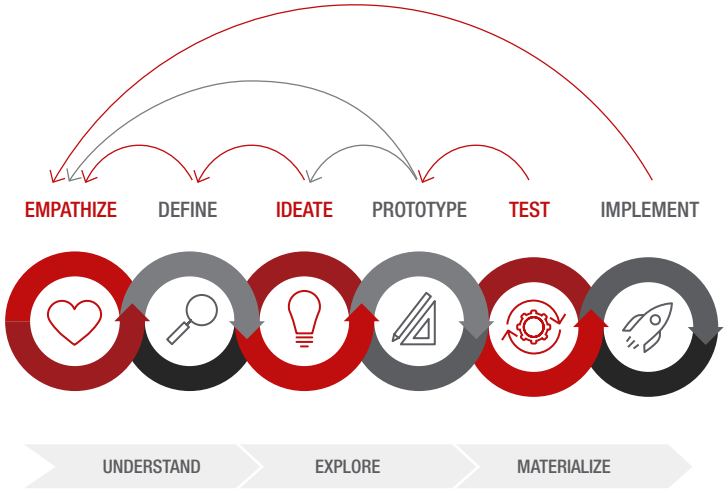
1. **It is not about how a product looks, but how a product works:** while the term 'design' can be reduced to the mere look and feel of a physical product, it is only one part of the design thinking methodology. Designing, in this case, combines the art of thinking about the functionality, features, usability, and the looks through the whole process [Berk (2017)].

2. **It is not about agile or scrum, but about solving hard problems:** design thinking is a human-centered solution-driven cognitive process from which design concepts (e.g., products or services) emerge. Agile, on the other hand, is an approach for software development under which requirements and solutions evolve through the collaborative effort of self-organizing and cross-functional teams [Cooper-Wright (2016)].

3. **It is not about execution, but inspiring ideation:** design thinking is about learning through iterations, being creative, and daring to prototype different solution approaches. The final design cannot be defined at the initial stages, hence a simple process execution to reach a specific goal is not a part of design thinking. Henry Ford once said, "If I asked what people wanted, they would say faster horses." So, challenge to innovate.

4. **It is not about the process, but about a mind shift:** design thinking stands and falls with the team's knowledge of the methodology to create a suitable solution through a research and invention process that has to be embedded within a company's culture. One cannot just read a book about it and execute design thinking following a construction kit [Kadam (2018)]

Figure 1: The iterative process of design thinking, where the status quo is constantly questioned leading to the best solution



Source: https://bit.ly/2CvEu1Z

Depending on the size of the task, different iterations and building steps can be used. However, they always embody the same principles laid out in the design thinking cornerstones below [Dam and Siang (2018)].

- Design thinking **starts with empathy**, being curious and conservative is key to start a human-centered design and keeping the final user in mind throughout the process.
- **Reframing the perceived problem or challenge** at hand and gaining perspectives, which allow for a more holistic look at the path towards these preferred situations.
- **Collaborative and multi-disciplinary teamwork** is endorsed to leverage the skills, personalities, and thinking styles of many in order to solve multifaceted problems.
- **Convergent styles of thinking assist to isolate potential solution streams**, combining and refining insights and more mature ideas, which pave a path forward.
- **Tests the prototypes** that survive the processes further to remove any potential issues.
- **Iterates** through the various stages, revisiting empathetic frames of mind, and then redefining the challenge as new knowledge and insight is gained along the way.
- It **starts off chaotic** and cloudy steamrolling towards points of clarity until a desirable, feasible, and viable solution emerges.

2. INHIBITING FACTORS TO CUSTOMER-CENTRIC INNOVATION

2.1 Product-centric organization

Banks have historically preferred organizing along client segments and products. A typical banking organization will divide competencies, especially in the front- and middle-office, along private and corporate clients, and across retail, investment banking, and asset management.

Further, the client offering is structured in product silos – payments, cards, investment products, and financing are all distinct product lines managed in separate divisions with dedicated product managers.

By itself, this can make a lot of sense – matrix-style organizations tend to be efficient in allowing for cross-pollination of information across functional domains and business lines. However, customer-centric innovation is a function of both increased probability of new ideas being permitted to germinate and an organizational setup conducive to executing on those same ideas.

It can be argued that banks today are inherently averse to innovation as they function in highly compartmentalized divisions, with fairly weak execution structures. Efforts are underway to establish 'digital factories' and 'innovation labs', however these initiatives are seldom embedded into the wider organization, tend to be expensive, and produce results that are often removed from the realities of daily business. A more promising approach would ensure service-oriented organizations, organized with customer needs and their respective service portfolios in mind.

2.2 Financial and technical constraints

Major financial services firms need to budget in comprehensive, annual cycles. This stands in stark juxtaposition to a 'fast-failure and early-success' mentality, which accepts uncertainty as part of the process, and cannot always accommodate the budgeting process used today. While funds are allocated, banks tend to require comprehensive business cases, requirements analyses, solution designs, vendor assessments, and

detailed resource calculations up front. Rarely is time and funding allocated appropriately for explorative causes, to identify customer needs, rapidly test new ideas, and establish a strong, customer-centric project proposal.

That is not to say that the rigor of change portfolios is unnecessary – it definitely remains so. However, the question remains whether innovation can be preplanned and budgeted. A more promising approach would entail launching projects into a time-boxed, explorative phase, and gradually requiring refinements around financials. This can still be accomplished with annual budgeting cycles and can help lower costs via an improved product-market fit, coupled with an agile delivery approach that aims to deliver iteratively.

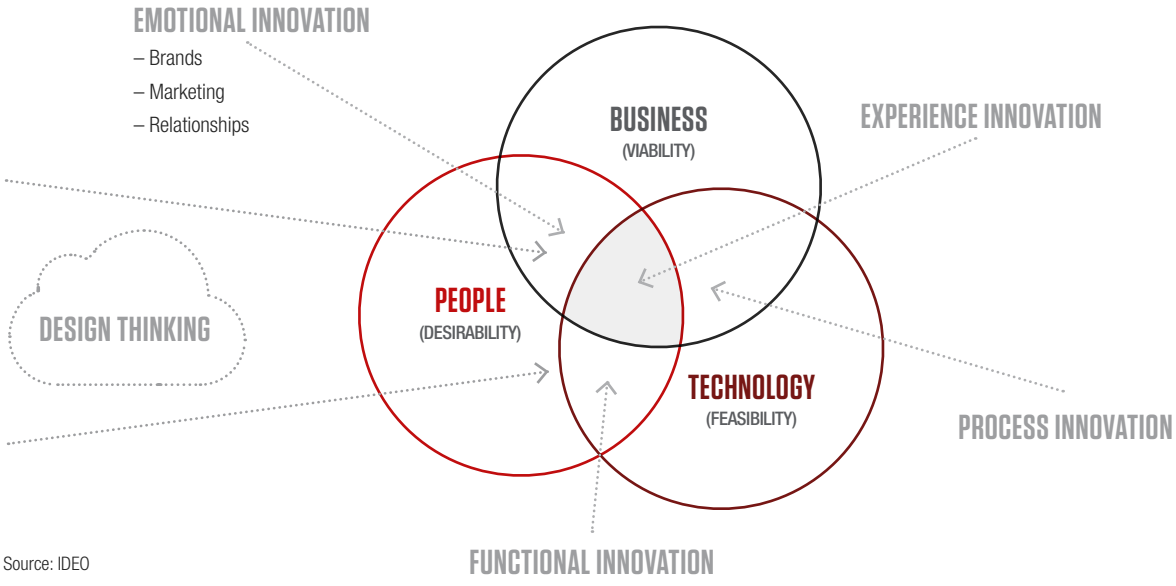
Further financial constraints are generated through legacy dependencies. Any effort to innovate on services often results in exponential investment in legacy platforms – thus making otherwise promising business cases unviable. Innovation requires an ability to leave behind the established status quo, and technology platforms in banks can pose the primary challenge to this effort.

2.3 Risk culture and risk aversion

Risk culture within financial institutions has evolved significantly since the global financial crisis. Led by American and European regulators, newly imposed rules on risk management and capital requirements have led financial institutions to divert resources to regulatory transformation and remediation, while investments in explorative undertakings such as research, development, and innovation initially plummeted. It took the financial services sector a few years to realize that holding back on innovation is not helping them move forward, so during the economic recovery most banks built incubation "labs" where even risk questions could be tackled.

The number of innovation labs have been increasing within the banking industry in recent years, and their numbers are growing. 87% of financial services firms say that they either have an innovation lab or have at least carved out some real estate for innovations.[1] This represents a 27% increase in the number of innovation centers in the past year.

Figure 2: Human-centered innovation



Source: IDEO

[1] https://bit.ly/2Ryhi7r

They can serve an important purpose in helping financial services firms pilot new technologies in a sandboxed environment prior to an enterprise wide implementation. This model is financially viable as banks can run an innovation center on a variable cost basis.

3. HOW DESIGN THINKING CAN ACCELERATE INNOVATION

3.1 Design thinking places the customer at the center

At its heart, design thinking is about being curious. It is about being a keen observer of things around you. You need to be curious about why things are the way they are, why things do not work, or why people behave the way they do. Once you nurture the mindset of being curious, you let go of judgment and seek to better understand everything around you [Kadam (2018)]. Being observant is about paying attention to the finer details. You need to ask questions when you start assuming and seek to understand what you do not know.

The next important thing that follows curiosity is empathy. When designing products, solutions, or business models for someone else, the biggest challenge is to understand the people you are designing for. And, therefore, the biggest mistake is to develop solutions without including the end-user, as it is often the case with banks. Often, the end-users' interactions are taken for granted, or worse, we tend to assume how they experience the world.

Banks' R&D departments and isolated project teams often disregard the issues that arise by not analyzing the needs of the end-user and solely developing on the basis of their knowledge and acceptance criteria. The golden key is to get an understanding of users' mental models, how the world looks from their perspective, and what their true needs are that need to be satisfied. Addressing the discovered needs not only satisfies the user, but also transforms an initial invention into a true innovation.

It should be noted that design thinking is not only for easy and functional "design" products, as all business relevant factors are taken in account throughout the development process. While developing empathy for the user and keeping them at the heart of the process is important, it is also crucial that the entire business perspective is kept in mind. To hit design thinking's sweet spot, the team has to not only consider the solution's human-psychological factors, but also ask whether it is desirable and, more importantly, viable and feasible [Brown (2009)].

3.2 The effectiveness of design thinking

Some might ask why the free-spirited design thinking approach should be chosen over other ideation methodologies. We hope that the following discussion can answer this question.

- **Risk reduction:** design thinking reduces the project and development risks through a continuous learning process. The so-called "fail forward" culture brings potential misperceptions from the beginning and throughout the process to light. Mistakes can (and have to) be made even in the early stages of the process to create a learning effect and steer the development in the right direction. Consequently, the project risks can be continuously reduced.
- **Failing forward:** the general high-end corporate setup has an unspoken zero-mistake-culture and thus **"failing" is often correlated with a negative outcome. Design thinking, however, paraphrases failure into learnings** and insights that are subsequently used to learn as an individual or team and ipso facto improve the ongoing development.
- **Fast cycles:** to achieve as many failures and win insights as possible, design thinking asks for fast and short iterative cycles. This **"fail often and therefore early"** is to be seen in coherence with "failing forward" and can happen through the process or at the end of a cycle [Leifer and Steiner (2010)]. The key is to set fixed development cycles and gather feedback from the end-user to spot the design problems; an approach that can lead to resistance in a waterfall dominated industry and, therefore, requires dedication from management and the whole team to follow the methodology. Fast testing and feedback cycles lead to quicker development and greater acceptance rate, which helps in the long-run.
- **The value of tangible prototypes:** all results in design thinking have to be in the form of tangible prototypes. Ideas are in general not real; hence prototypes can make ideas real, tangible, and testable [Brown (2009)]. While the demand for a physical prototype of a technical system is understandable, it can seem absurd to create a tangible prototype for a service-based business. A variety of successfully completed projects have shown that the use of storytelling, for example, can create the sense of tangibility. Furthermore, the use of prototypes has shown the benefit of simplifying complex problems by separating the overall target product into smaller components. This allows for a more focused development in the team [Brenner and Uebernickel (2016)].

In general, design thinking has the ability to reduce a project's risk factors, raise the customer's acceptance rate, and even deliver desirable solutions more quickly. This all is possible if the methodology is well thought through and accepted by all of the stakeholders.

> "If banks can develop a design thinking friendly environment and recognize the method as a promising means to foster innovation, they stand a fighting chance."

4. BEHAVIORAL CHANGE TO DEPLOY DESIGN THINKING

Taking into consideration the various factors that drive successful application of design thinking, we propose the following actionable behavioral changes that banks need to pursue.

4.1 Space, absence of fixed processes, and allowing change

Physical space and the work environment have emerged as key factors to facilitate change. Through adapting the physical environment, organizations are able to lower hierarchical boundaries, enhance ideation and creativity, foster and accelerate prototyping, and increase the rate of learning and change. The key concept for the spatial setup is flexibility. Space ought to allow for and support any kind of ideation and prototyping activities. Going through a number of rapid iterations, testing ideas, and the boundaries of the solution space via prototypes, allows the project teams to increase their rate of learning significantly.

Research points to the inability of any particular fixed process model to support the output of radical new products and services. A great example is DARPA, the U.S. Defense Advanced Research Projects Agency that has supported groundbreaking projects ever since its establishment in 1958. Unlike the majority of other business, government, or academic research organizations, DARPA is specifically and solely focusing on the creation of radical engineering and system innovations. Their mission is to "maintain the technological superiority of the U.S. military and prevent technological surprise from harming our national security by sponsoring revolutionary, high-payoff research bridging the gap between fundamental discoveries and their military use" [DARPA (2010)]. In fact, all projects may be defined by a set of criteria that have found their way into the urban directory as 'DARPA hard' [Urban Dictionary (2011); Van Atta et al. (2003)].

All projects must be:

a) Technically challenging (beyond current limits),

b) Actionable (proof of concept or prototype),

c) Multidisciplinary (complex), and

d) Far reaching (advances on a grand scale, radical).

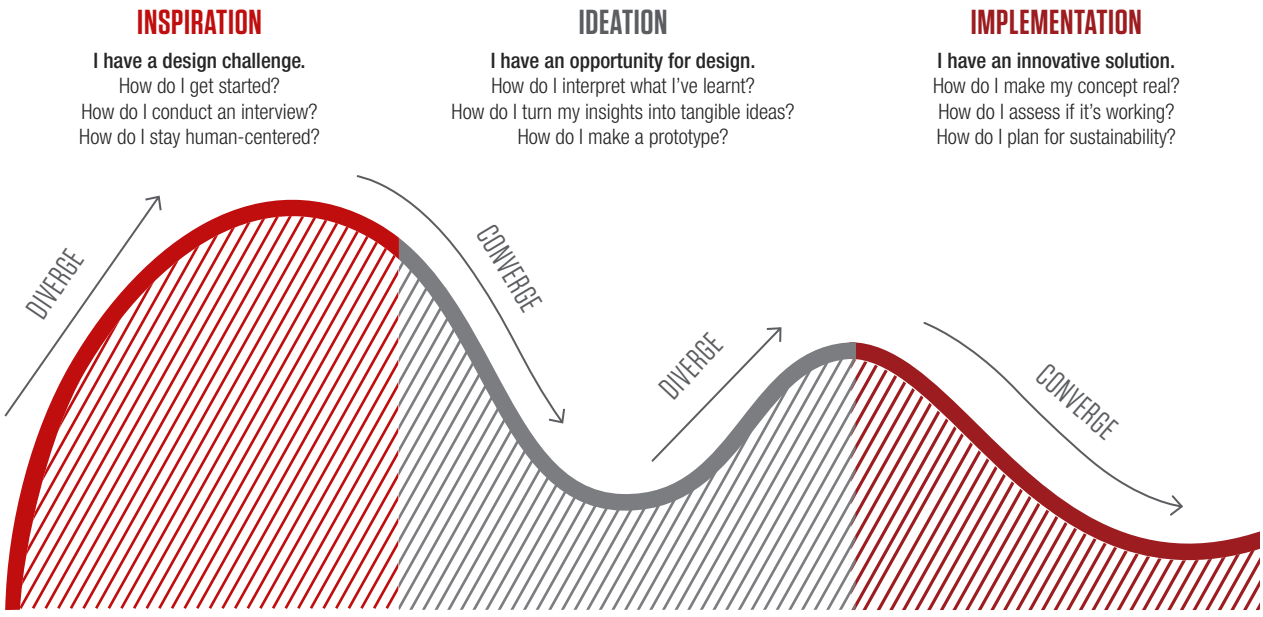
Carleton (2011) has securitized this remarkable test bed in her doctoral thesis. The most interesting observation concerning change is when projects and people are vision driven, have a focus on workshops and prototyping, and disregard documentation and project management rules.

The selection of DARPA programs depends on the creation of a vision. It is the starting point for any program, and the project champion, the program manager, embodies it. The ideation and iteration of this vision serves as the central focal point for the usually dispersed subprojects, teams, and stakeholders. Envisioning a certain technological future does not define or limit the future projects; it serves as an indictor of the current direction of the organization's efforts.

Interestingly, the main instruments to generate, iterate, and re-formulate such a vision are workshops and the creation of proof of concepts or prototypes at various stages. The first allows the socialization and evolution of the visionary ideas amongst all participating stakeholders, while the latter allows tangibly communicating and even testing the vision at various critical junctures. The program and project managers also enjoy a remarkable freedom from established processes and rules. No established system or documentation requirement is forced onto their activities. Prototyping is the norm and the specific activities follow the actual demand of the specific task at hand.

No institutional models are limiting people and their behaviors for the sake of generating economies of scale. Innovation and change are the generation of the new – the primary goal is the best outcome at certain budget constraints, not its process efficiency in terms of minimal resource allocation. Another point to consider is

Figure 3: The design process



Source: IDEO

the ways that go and no-go decisions are prepared and executed. Instead of relying on peer review processes or committees or other group-based decision tools (not to speak of pseudo quantitative stage gate filter variables), decisions are taken by the leaders who ultimately bear the responsibility for success or failure. Failure is accepted and preplanned. The underlying rationale is that peer review and committees are in fact hedging mechanism for taking tough decisions at the extreme end of the possible solution spectrum. They will inherently favor outcomes close to the sample's median opinion. Hence, traditional decision tools would prevent DARPA from actually attempting to deliver radical innovations. All of these activities, and most importantly the absence of fixed processes, serve to generate change that comply with the idea of DARPA, as described above.

As the prior discussion on space and flexibility, as well as on the absence of fixed processes, indicated, the major concept to support change and learning, the generation of radical new product and system solutions, is to allow change to happen. We do not assume to have control over the existing solution space, so we cannot preplan or manage it. Indeed, we are aware that the concepts that challenge the established dogmas have a higher chance to deliver radical improved value. Any systematic and fixed support system, inhibiting the creative use of space and the employing and combining of new processes, seems to counter the notion of change. Consequently, we attempt to provide the physical, organizational, procedural, and mental environment that allows the project teams to experiment and to prototype. This becomes especially difficult when proposed solutions counter the experiences and knowledge models of the professors and coaches.

Instead of prematurely ending the iteration processes at this point, we allow, indeed support, the testing of theses, ideas, and concepts. Very often, a failed prototype test, the hitting of the boundary of the possible solution space, generates the winning insights for either an extreme solution along that line or, even better, a new way that allows circumventing the existing limitations. To generate this kind of change, we attempt to minimize institutional, organizational, and procedural boundaries. We emphasize and support flexibility, and we force ourselves to let change happen. Hence, we do not prescribe procedural recipes. The focus on people and team development should be on skills, moves, and the demand for tangible prototypes. A word of caution: allowing, and even fostering of, this kind of ambiguity is difficult and demanding for the coaches and requires a conscious effort, especially on behalf of individuals who have to unlearn their organization skills to a certain extent.

4.2 Design process of divergence and convergence

Contrary to the classical and analytical design process applied for the development of incremental changes, the design process aiming to radical changes can be seen as an iteration of divergent and convergent activities [Alexander (1964)]. Banathy (1996) describes the divergent activity as "...consider a number of inquiry boundaries, a number of major design options, and sets of core values and core ideas. Then we converge, as we make choices and create an image of the future system."

This divergent-convergent process may be depicted as slowly closing funnel, linear over time [Cross (2000); Ulrich and Eppinger (2008)], or as repeating design cycles, spiral like, that iterate through the generic prototyping phases of design, build, and test [Thomke and Fujimoto (2000)]. The classical convergent phase is about optimizing the answer. It is deductive and inductive in nature and may comprise simple tools, such as the Pugh Chart [Pugh (1996)] or a Quality Function Deployment [Hauser and Clausing (1988)], or run on complex model simulations and optimizations.

The design thinking approach that we are favoring not only emphasizes the circular or spiral nature of the process (feedback loops were common but limited in the classical process models), but it clearly identifies the need of divergent search activities. Developers are constantly and rapidly going through design-build-test cycles. In each cycle, during the divergent phases, we are focusing on the problem rather than on the solution, trying to understand who really is the user, which elements are truly involved, how many other ways are there to solve the problem, and can we rephrase the challenge and circumvent the problem? These divergent activities usually result in a number of ideas or concepts that are in a next step built and then down-selected by testing.

4.3 Rapid and tangible prototyping

In the design process with clients, we concentrate on creating prototypes as fast as possible in order to test particular ideas, the design hypotheses behind the prototype. Speed of learning is key. As a result, our prototypes tend to be of low resolution and physical or tangible rather than virtual. Depending on the design stage, whiteboard, simple cardboard, and duct tape constructions, prototypes made from wood or clay etc. might be created. Each prototype is built to test a specific idea and/or a system interaction. They range from simplistic rough artifacts that merely resemble an idea (communication prototype), to lookalike prototypes (conveying certain external property ideas), to critical functional and functional prototypes (technical proof of concepts), to alpha and beta prototypes.

It must be noted that later stage prototypes cost an order of magnitude more in resources, both in time and money than early prototypes. It is, therefore, essential to concentrate on the early stage or fuzzy front end of the new product design. One of the most pertinent recent insights, based on dissertational related work of Jonathan Edelman [Edelman et al. (2009)], is that the choice of the prototype material or environment directly influences the amount and degree of the generated alternatives. The breadth and depth of the solution space explored seems to relate to the sophistication or resolution of the prototyping materials employed. A sophisticated CAD prototype is least likely to be considerably changed in following iteration cycles. The product architecture is implicitly fixed and the software and its capability limits possible ideation changes. Tangible 3D prototypes allow the creation of more alternatives with relative ease. These types of lookalike prototypes are especially good in conveying ideas and form factors to non-specialist users. However, once this level of resolution has been reached, changes tend to be incremental. If we contrast this to using very basic prototype material, simple cardboard, or even just a sketch, the possibility for more radical and faster iterations, and thus learning, is obvious.

As a rule of thumb, the early stage product development determines the level of radicalness of the final solutions. We, therefore, advise product development teams to stay in this early phase for more than a third of the entire available project time. We have to force ourselves to abstain from entering solution optimization in order to gain intimate awareness of the problem space. This increases the chances of us generating the real breakthrough idea we are looking for.

4.4 Need-finding, user testing, and experience enactment

Central at the early stage of the new product or system concept design is an intimate understanding of firstly, who actually is the end-user, and secondly, what are the real user's needs that we aim to satisfy with the solution. Often, projects start with a fixed set of specifications and requirements. This approach, very suitable for incremental change and innovation, focuses the attention and resources onto the optimization and execution of the selected concept. Time and again, final solutions do not meet end-user needs and need to be re-designed before deployment can succeed. This costs significantly more money and prestige than conducting more exploration early on. Consequently, we are concentrating on the first phase of the design process.

The first challenge lies in identifying the end-user to design for. Some iterations and perspective changes may unearth surprising users. Challenged to redesign satellite architecture, a research team at Stanford is currently focusing on the testing engineer as the target user. The pre-launch testing process ties significant resources due to the fact that satellites have not been engineered for modularity/mass customization and access to the sub systems that have to be tested again is typically outside the satellite design team's concern. Testing and validation become a large fraction of system integration costs that are, in turn, a major factor in net deployment cost.

In an unrelated case study, scrutinizing medical device development, it is not the patient, the obvious user, and their needs that are central for the success of a new product. Though, any new solution must at least be equal in terms of patient value added, the real litmus test lies in the value gained by the hospital and insurance companies, in relation to the change required by the practicing medical doctors. Who is the user for whom we have to design for in this case [Aquino et al. (2011)]? Once a single target user or a user system was identified, the researches attempted to gather information on the underlying needs that ought to be satisfied by the new solution. While surveying and interviewing users does give valuable information, very often users are themselves not capable of expressing their needs. Indeed, when confronted with something absolutely new, for example a device based on a new technology or material, users can only draw from analogies and not answer from experience. Even if they can, very often their personal perspectives are too limited to truly understand the problem. Observations, especially when analyzed systematically using video interaction analysis, result in a better understanding of the process and behavior we intend to improve. As the literature of knowledge management tells us, this direct tactile involvement with the problem is often the only way to transfer implicit procedural knowledge. As Nonaka and Takeuchi (1995) describe, to build a home bread-baking kitchen equipment that also kneads the dough, it was necessary to practice kneading with a baker. The development team would not have been able to uncover the complexity of the compress, pull, and twist action necessary to create dough that rises just right.

5. CONCLUSION

Banks have great challenges, in both method and culture, to overcome before design thinking can be deployed successfully. If banks can develop a design thinking friendly environment and recognize the method as a promising means to foster innovation, they stand a fighting chance. Design thinking requires allocating time and resources that may or may not have direct, measurable impact on top and bottom lines. The methods do not focus on execution and have little or nothing to do with agility in delivery. Too often, banks conflate agile transformation, technology, and design thinking, while the latter is a problem-solving technique that sits at the intersection of business, technology, and humans. This article intended to show these differences and point out that the path forward can be as simple as creating the right physical environment, introducing a rapid prototyping approach (meaning prototyping an idea within hours, not even days, using basic materials), and involving the end-user in the process from the start. Design thinking in financial services holds a lot of promise. It remains up to banks to harness its power.

REFERENCES

Alexander, C., 1964, Notes on the synthesis of form, Harvard University Press

Aquino, S. L., M. Steinert, and L. J. Leifer, 2011, "Designing to maximize value for multiple stakeholders: a challenge to med-tech innovation," Working Paper

Banathy B. H., 1996, Designing social systems in a changing world, Plenum

Berk, S., 2017, "Design thinking is not thinking like a designer, it's channeling personas," Medium, February 22, https://bit.ly/2sWzzmi

Brenner, W., and F. Uebernickel, 2016, Design thinking for innovation: research and practice, Springer

Brown, T., 2009, Change by design: how design thinking transforms organizations and inspires innovation, HarperBusiness

Carleton, T. L. (2011). The Value of Vision In Radical Technological Innovation. Stanford University, Stanford.

Cooper-Wright, M., 2016, "The blurring between design thinking and agile," Medium, November 24, https://bit.ly/2Hn4c7F

Cross, N. (2000). Engineering design methods: strategies for product design (Vol. 58). Wiley Chichester.

Dam, R., and T. Siang, 2018, "Five stages in design thinking process," Interaction Design Foundation. https://bit.ly/2eBn1uh

DARPA (2010). Defense Advanced Research Projects Agency. DARPA Mission. Retrieved from http://www.darpa.mil/mission.html.

Edelman, J., & Currano, R. (2011). Re-representation: Affordances of Shared Models in Team-Based Design. Design Thinking, 61-79.

Edelman, J. A., Banerjee, B., Jung, M., Sonalkar, N. and Lande, M. (2009). Hidden in Plain Sight: affordances of shared models in team-based design. Proceedings of the International Conference on Engineering Design. Presented at the ICED 2009, Stanford.

Eris, O. and Leifer, L. (2003). Faciliating Product Development Knowledge Acquisition: Interaction between the Expert and the Team. International Journal of Engineering Education, 19(1), 142-152.

Gottman, J. M. and Levenson, R. W. (2000). The Timing of Divorce: Predicting When a Couple Will Divorce Over a 14 Year Period. Journal of Marriage and Family, 62(3), 737-745.

Hauser, J. R. and Clausing, D. (1988). The house of quality. Harvard business review, 66(3), 63-73.

Ju, W. G. (2008). The design of implicit interactions. Stanford University.

Jung, M., Chong, J. and Leifer, L. (2010). Pair Programming Performance: An emotional dynamicspoint of view from marital pair counseling. Electronic Colloquium on Computational Complexity ECDTR, (2)

Kadam, A. R., 2018, "Design thinking is not a process, it's a mindset," Entrepreneur," March 12, https://bit.ly/2Mnw29n

Leifer, L., and M. Steiner, 2001, "Dancing with ambiquity: causality behavior, design thinking, and tripleloop learning," Information Knowledge Systems Management 10, 151–173

MacCormack, A. (2001). How internet companies build software. MIT Sloan Management Review, 42(2), 75-84.

ME310. (2011, January 30). ME310. Retrieved from http://www.stanford.edu/group/me310/me310 2010/about.html.

ME310 EXPE 2011 (class report). (2011).

Nonaka, I. and Takeuchi, H. (1995). The knowledge-creating company. New York, 1, 995.

PRL (2011). Product Realization Laboratory. Retrieved from http://www.stanford.edu/group/prl/prl site/.

Pugh, S. (1996). Concept Selection–A Method that Works, Chapter 14 from "Creating Innovative Products using Total Design," Addison-Wesley.

Sirkin, D. M., Sonalkar, N., Jung, M., & Leifer, L. J. (2009). Lowering Barriers to Distributed Design Research Collaboration. Proceedings of the 17th International Conference on Engineering Design (ICED'09), Vol. 9 (pp. 279-286).

Tang, J. C. (1989). Toward an understanding of the use of shared workspaces by design teams. Stanford University.

Tang, J. C. (1991). Findings from observational studies of collaborative work. International Journal of Man-machine studies, 34(2), 143-160.

Tang, J. C. and Leifer, L. J. (1991). An observational methodology for studying group design activity. Research in engineering design, 2(4), 209-219.

Thomke, S. and Fujimoto, T. (2000). The effect of front-loading problem-solving on product development performance. Journal of Product Innovation Management, 17(2), 128-142.

Ulrich, K. T. and Eppinger, S. D. (2008). Product design and development. McGraw-Hill Higher Education. Retrieved from http://books.google.com/books?id=WTNoSgAACAAJ.

UrbanDictionary. (2011). UrbanDictionary–DARPAhard. RetrievedfromhttP://www.urbandictionary.com/define.php?term= DARPA%20hard.

Van Atta, R. H., Cook, A., Gutmanis, I., Lippitz, M. J., Lupo, J. and VA, I. F. D. A. A. (2003). Transformation and Transition: DARPA's Role in Fostering an Emerging Revolution in Military Affairs. Volume 2-Detailed Assessments.

UNDERSTANDING THE VALUE OF DESIGN THINKING TO INNOVATION IN BANKING[1]

CLAUDE DIDERICH | Managing Director, innovate.d llc

ABSTRACT

With the advent of fintech, the banking world has been confronted with the method of design thinking, a proven method for solving wicked problems. Design thinking unleashes creativity and supports developing innovative solutions that are desirable (customers are interested in buying), feasible (banks can deliver upon the promises made), and viable (banks can make a profit). It puts the customer center-stage and focuses on satisfying customer needs and understanding customer jobs-to-be-done. Through its iterative approach, design thinking delivers differentiated and superior solutions, both from a functional and an emotional perspective. By observing customers in their natural environment, prototyping and validating ideas, design thinking ensures that developed solutions work. This article discusses how design thinking can aid in making banking more innovative.

1. INTRODUCTION

Many banks have understood over the recent years, specifically since the advent of fintech, that innovation is necessary for success. It is not uncommon for larger institutions to have named a Chief Innovation Officer or launched diverse innovation projects. Unfortunately, most of these initiatives have yet to unleash their full potential. A mistake often made is to assume that innovation predominantly belongs to IT. Another reason for their limited success is that banks tend to foster innovation inwards-out, focusing on business process improvements, cost reducing digitization, or product engineering. A third reason for the lack of accomplishments can be found in the inherent business model of banks – intermediating financial assets in a highly regulated environment, like cash deposits to loans, equities to investments, or payments. Indeed, exploring regulations to come up with creative ideas that offer added value to the parties involved in intermediation is a wicked challenge.

If you have ever been involved in an initiative to develop and launch a new product or service, the following characterization probably sounds familiar to you. Gyro is a bright banking employee who has come up with a groundbreaking idea for a mortgage product where customers can dynamically adjust the interest rate exposure, rather than having to wait for the loan to mature. He started by preparing a PowerPoint presentation describing the idea and presented it to his manager. As the idea was not killed right away, he presented it to multiple committees, each time embellishing the PowerPoint presentation in a different way to meet the

target audience's preferences. Throughout this process, his idea got watered down and drifted away from his initial vision, without really improving upon it. After he finally succeeded in convincing all those committee experts and getting a business case approved, he no longer deemed it necessary to seek formal feedback from customers. Then, with limited involvement of Gyro, IT launched a project to develop the supporting systems. Typical waterfall project management methods came into play. After double the time planned, and significant budget overruns, the new mortgage product was finally launched, just to find out that customers didn't understand it and failed to see the added-value in it.

"If traditional banks want to survive in the ever faster changing environment, they must become better at defining and exploiting their competitive advantages."

At least since the latest financial crisis, how banks approach innovation has come under pressure from fintech startups, like Betterment, Revolut, or LendingClub, as well as large non-financial players, like Apple, Amazon, or Alibaba. These competitors exhibit superiority in four key areas:

- They do not have to worry about legacy systems and are therefore more agile.
- They own superior capabilities in exploiting economies of scale.
- They focus on addressing specific customer jobs in a superior way, rather than trying to offer everything for everyone.
- They apply a distinct method to problem solving, leaving linear, business case-oriented planning approaches behind, and focus on agility.

When analyzing why banks have such a hard time competing in the innovation space, three categories of root causes can be identified:

1. Technological: many banks still rely on legacy IT systems to support their core banking operations. The technology know-how of these platforms is hidden in an oversized IT division, resulting in senior management not getting the full clarity and transparency of information required for efficient decision-making. This leads to cumbersome IT projects as well as a lack of business focus and customer-centric use of technology.

2. Cultural: even more prominent than technological aspects, are cultural reasons deeply rooted in the banker's DNA, hindering innovation. Typically, banks are risk averse, which reflects itself in a change-averse culture. The argument "but it worked fine in the past" predominates. Another key objector is the existing silo mentality and the "not invented here" syndrome. Finally, the cultural belief that banks know what the customers need better than they do results in not-so innovative solutions and not-so meeting real customer needs.

3. Business model: the third category of reasons that hinder innovation in banking are their approaches to doing business. The business model of a typical bank has not changed over the years. Banks are very slow to embrace new trends, hoping they pass by, as business cycles do. Limited number of new to the market products have been introduced. In addition, the market driven characteristics result in banks floating on the waves rather than riding them.

If traditional banks want to survive in the ever faster changing environment, they must become better at defining and exploiting their competitive advantages. The most prominent competitive advantage that banks have over startups is that they own trusted customer relationships. Indeed, various fintech firms have learned the hard way that acquiring new banking customers is much harder than acquiring typical consumer business customers.

But no customer relationship lasts forever. Banks must re-learn nurturing their customers' trust by focusing on

- solving real customer problems, rather than selling off-the-shelf products,
- offering a compelling customer experience that fosters trust, and
- focusing on delivering value for money, as perceived by customers, rather than the bank.

Being successful in banking requires putting the customers center-stage and supporting them in getting their jobs done [Christensen et al. (2016)].

2. DESIGN THINKING

In recent years, design thinking has become a valuable method for solving wicked problems. But what is design thinking, and why does it address the challenges faced by today's banking industry in a superior way?

Design thinking is a human-centered, iterative method for creative problem solving that draws from the designers' and architects' toolkits by integrating:

a. the needs of people, including customers – ensuring **desirability**,
b. the possibilities of available capabilities, including technologies – ensuring **feasibility**, and
c. the requirements for business success, that is, profitability – ensuring **viability** of the solution. [derived from Brown (2009) and Kelley (2001)]

Design thinking takes a different approach to looking at the world. It focuses on "doing in a collaborative way," rather than "planning in corner offices." By learning from creative people, design thinking focuses on developing and improving solutions in an incremental and iterative way. At the core of design thinking stands **abductive reasoning,** starting with a set of abstractions and seeking for the simplest and most likely solution. The initial solution is then improved upon through inference towards a great solution. Unlike deductive reasoning, abductive reasoning does not assume that the solution is contained in the premises of the problem. It is based on Einstein's saying, "we cannot solve our problems with the same thinking we used when we created them".

But design thinking is more than just a problem-solving method, it is a **problem-solving ecosystem fostering innovation**. This ecosystem, which defines the design thinking culture, is made up of three key characteristics:

1. A design team, and its members, exhibiting diverse traits and bringing varied expertise and experience to the table.

2. A location where the team can be creative and thrive, sometimes a garage, a loft, or a lab.

3. A method and associated frameworks supporting the creative process by giving it structure, focusing on combining divergent and convergent thinking.

2.1 Teams

Design thinking is based more than any other creativity or problem-solving method on teamwork. Putting together a great design team is a challenge, the first of many faced on the journey to success. Team members should cover the diverse skillsets that innovation requires and include visionaries (rising above the status quo), troubleshooters (fixing short-term problems), iconoclasts (challenging the status quo on any occasion), pulse takers (those who can obtain the perspectives of stakeholders through formal and informal channels), craftsmen (offering expertise in building and prototyping), technologists (functioning as subject matter experts), entrepreneurs (great in thinking and doing but needing freedom to thrive), and cross-dressers (enthusiasts that are always open for something new) [Kelley (2001)]. Teambuilding is about leveraging diversity, sameness is not the goal.

Just assembling the right people, building a team with a great spirit, is not enough. To be successful in design thinking, team members must be fully committed. How often have you been in a team meeting when a key player left half-way though because they had a different, more important, meeting to attend. In design thinking, this must not happen. Solving the challenge at hand must be the design team members' top priority. Design thinking is not a part-time job. And it is especially not a bandwagon on which one can hop-on and hop-off. Design team members must be engaged throughout the whole problem-solving process. This does not mean that anyone involved must be fully committed. It only means that those not fully committed cannot be part of the core design team. Their role reverts to one of stakeholder, like a customer, a risk manager, or a back-office employee, providing an opinion. In design thinking, they are called informants.

2.2 Location

A team, by its mere definition, requires individuals working together toward a common goal. This implies collaboration. Although today's technologies allow collaborating from remote locations, successful design teams interact physically, most of the time. Technologies, like Skype, Conceptboard, or Google Hangouts, are great for interacting with informants, but not for creative problem solving.

In addition to being present in the same location, creativity needs a unique working environment. This environment must allow designers to interact and brainstorm, as well as provide quiet places to think. This does not mean that every bank needs to transform their offices into a Google-like playground. It means that windowless cubicles do not work. Depending on the challenge to be addressed, different working environments are most appropriate. Sometimes a large meeting room with blank walls and lots of flip-charts and post-its will do. Sometimes more evolved setups are needed. Ideally, the design team members should be able to configure their own working environment. Banks should view offering their creative minds a compelling work environment as an investment, rather than an expense. Combined with the right team, it will pay off multiple times through increased productivity. According to Kelley (2001), creating a great working environment is nearly as important as hiring the right people. Both are indispensable.

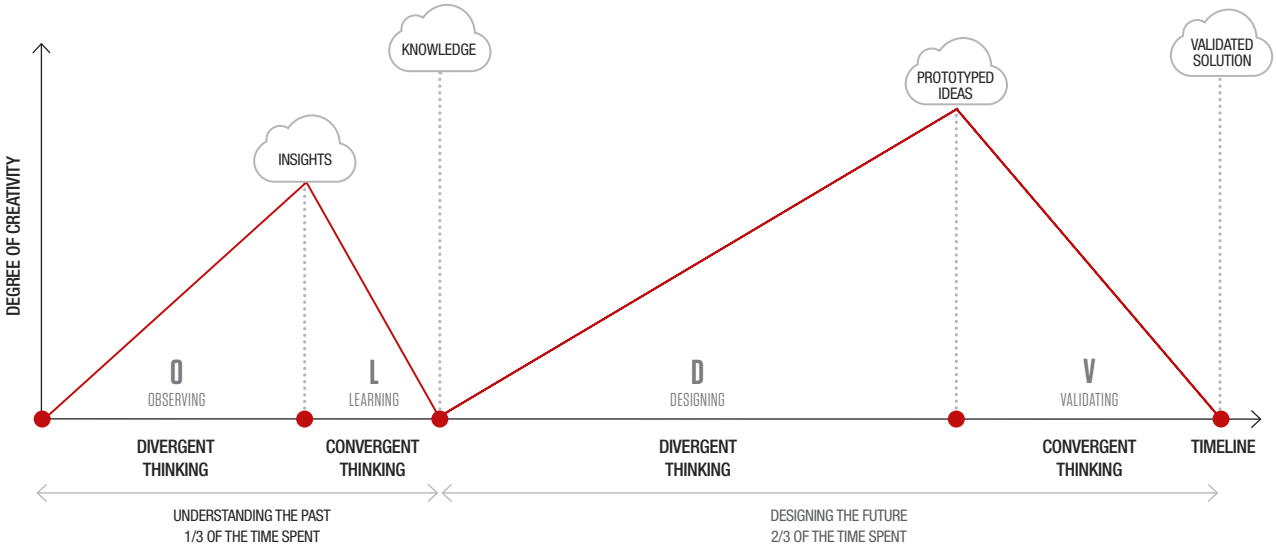
2.3 Method

The design thinking method, which finds its roots in Simon (1968), is based on iteratively combing phases of divergent and convergent thinking, working towards a feasible solution. Work by Arnheim (1969), McKim (1973), Lawson (1980), and Cross (1982, 2011), amongst others, refined the method over the years. A diverse set of design thinking method variations have emerged, from e.g., Stanford's d.School, University of Virginia's Darden School of Business, the MIT Sloan School, the Hasso Plattner Institute of the University of Potsdam, or the University of St. Gallen. Although each of these variations of design thinking includes different terminologies and sometimes different steps, they all follow the same philosophy, that is, observe to learn, prototype ideas, and validate designs with real users. The design thinking method illustrated in this paper is based on the double diamond approach of the British Design Council. Figure 1 illustrates the four steps of the design thinking method, primarily supporting service design. Each of the four steps i) observing, ii) learning, iii) designing, and iv) validating, focuses on a specific outcome and builds upon the findings from the previous steps. If the input at any step is insufficient or inappropriate, the design thinking method iterates to fill the identified gaps.

2.3.1 OBSERVING (DIVERGENT THINKING, FOCUSING ON UNDERSTANDING THE PAST)

Observing is where design thinking starts. It focuses on objective fact-finding. Its goal is understanding the challenge to be addressed and screening the solution space from different perspectives. Observing aims at gaining a comprehensive understanding of the environment in which to design a solution, including identifying constraints and opportunities. This means that typical, as well as extreme, informants, those that have a strong positive or negative bias towards the challenge, should be observed.

Figure 1: Linearized representation of the design thinking method based on the double diamond approach



In contrast to the typical problem-solving techniques, design thinking does not start with asking questions and interviewing informants. Indeed, the so-called Henry Ford trap[2] must be avoided. Design thinking is based on passive observation of informants, primarily customers, in their natural environment. No a priori root cause or potential solution should be assumed. Successful observation proceeds iteratively in a top-down manner, focusing on both functional and emotional aspects, which make up the actual insights universe. The focus is put on those observed insights that are deemed most relevant. Techniques developed in ethnography [Spradley (1980)] come into play.

Typically, observing wealth management advisory clients would involve focusing on what they are doing with the investment advice received. Are they reading the reports received or are they only browsing through them? Who do they involve in decision taking? How do they translate the advice received in a quantified transaction? What kind of feedback do they seek from the bank's advisor before trading? Observations to find answers to these questions and more are important for designing an advisory offering that best serves the customer's way of addressing their job-to-be-done, that is, investing. Facts matter more than opinions!

To complete the picture obtained from passive observation, exploratory interviews are conducted [Spradley (1979)]. Good exploratory interviewers spend about 20% of their time asking questions and 80% listening to the informants. At the end of the observing step, the design team has collected many objective insights around the challenge at hand.

2.3.2 LEARNING (CONVERGENT THINKING, FOCUSING ON UNDERSTANDING THE PAST)

Learning in design thinking means structuring the insights gained to obtain valuable knowledge that can be used as the basis for ideating, prototyping, and designing innovative solutions. Learning is about extracting knowledge in an agile way. Mastering the learning step is one of the secret ingredients of successful design thinking.

Learning starts with selecting one or more frameworks to structure the insights gained. Typical frameworks that come to application are personas [So and Joo (2017)], the customer journey map [Liedtka et al. (2014)], the

Figure 2: Summarized customer journey map derived from observing the persona Jenny, a new customer, while opening a bank account

DESCRIPTION	Identifying a bank	Visiting the selected bank	Getting advice about the offering	Completing paperwork	Returning home
RESPONSIBLE STAKEHOLDER	Jenny	Jenny	Bank advisor	Jenny	Jenny
INVOLVED STAKEHOLDERS	Friends and family	Bank support personnel	Jenny	Bank advisor	
FUNCTIONAL INSIGHTS	• Searches for banks on the internet • Asks friends and family members for suggestions	• Checks opening hours • Drives to bank • Looks for a parking spot	• Waits for the next available advisor • Listens to the advisor explain the offering • Ask questions • Looks at brochures handed over	• Decides which account offering to chose • Scans documents • Asks additional questions • Signs documents • Asks for copy of documents	• Returns home by car • Informs her employer of the new bank account to be used for salary payments • Waits for the credit card to be delivered by mail
EMOTIONAL INSIGHTS	• Is unsure about the bank chosen • Questions the trust in the bank identified	• Does not know what time is best to visit the bank branch to minimize wait time • Assesses bank based on greeting received	• Is somewhat lost with the large number of options • Feels reluctant to move forward because of a lack in understanding	• Feels pressured to sign documents • Does not understand the legal writings	• Feels relieved to finally have a new bank account

[2] Henry ford is often quoted as saying "If I had asked what customers want, they would have said faster horses. And we would never have invented the car." Although there is no factual evidence that Ford actually said this, history indicates that he was most likely thinking along these lines; that is, believing in the apparent inability of customers to formally state their unmet needs.

business model canvas [Osterwalder and Pigneur (2010)], and the value chain [Porter (1985)]. Frameworks help categorize insights, as well as separate the relevant from the irrelevant. Figure 2 illustrates a summarized version of the customer journey map used to describe opening a new bank account. As can be seen, the customer journey starts the moment they identify the job-to-be-done, that is, wanting to open an account at a new bank, looking for a bank branch to visit. Important knowledge, like how the persona Jenny assesses trust when arriving at the bank branch – namely through how she is greeted, rather than focusing on the building appearance or furniture – is identified. Another important learning documented in the customer journey map is that Jenney is having a hard time with the legal documents she must sign, providing a design opportunity to improve upon.

> "Just because a prototype looks promising to its designers does not mean that it will be accepted by customers."

2.3.3 DESIGNING (DIVERGENT THINKING, FOCUSING ON DESIGNING THE FUTURE)

Designing is where creativity is unleashed. Based on the knowledge gained so far, novel ideas or novel combinations of existing ideas, are generated. The popular ideation method, brainstorming [Osborn (1963)], is mostly used for ideation. Alternatively, more recent and more elaborated methods, like anticonventional thinking [Baumgartner (2015)], collaborative structure enquiry [Baer et al. (2013)], or the what-if-wall method [(van der Pijl et al. (2016)], may be applied. There is no single best ideation approach. Whichever method is used, design thinkers must avoid the trap of falling in love with their first idea.

During the second part of designing, ideas are transformed into solution prototypes. Prototypes may be physical solutions, mock-ups, conceptual illustrations, sketches, role plays, or even mental models. Prototypes do not have to be complete. Optional features should be left out. The only requirement that any prototype must fulfill is that it is sufficiently realistic to allow for testing its validity. Prototypes must allow the users to find out what works and what does not. Prototyping should follow the basic principles – keep it simple and focus on the essential. It is not uncommon, it is even typical, to iterate between ideation and prototype building. The LEGO® SERIOUS PLAY® method [Kristiansen & Rasmussen (2014)] allows for combining ideation and prototyping in a 3-D world.

Storyboarding may be used to describe how a customer can open a bank account purely online. Each illustration would focus on one process step, like the customer showing their passport or ID card to the webcam allowing the bank employee to identify them, or the customer signing documents with their finger on their mobile phones.

2.3.4 VALIDATING (CONVERGENT THINKING, FOCUSING ON DESIGNING THE FUTURE)

Just because a prototype looks promising to its designers does not mean that it will be accepted by customers. The validation step is a key feature of the design thinking method, not found in other problem-solving approaches. Not only does validation ensure desirability, feasibility, and viability of the prototyped solution, it also helps remove grid-lock discussions, often encountered in boardrooms [Liedtka et al. (2017)], by exposing decision making to field experiments.

Validation starts by formulating assumptions underlying the developed prototypes. Assumptions are prioritized based on their relevance for success and their complexity to validate. Validation experiments are designed and executed. The outcomes are used to improve upon the designed prototypes. In the context of business model prototypes, iteratively improving prototypes is called pivoting [Ries (2011)]. Validation is as much about learning from failure in a controlled way, as it is about mitigating risks. It is important to understand that validation in design thinking is different from hypothesis testing in statistics. Validation is forward-looking based on experiments involving actual and potential customers, whereas statistical testing is backward-looking and reliant on historical data. In addition, the goal is different. Validation is about supporting decision making, whereas statistical testing is about t- and p-thresholds.

Figure 3 illustrates a typical experiment card used to document a design thinking experiment, focusing on validating the assumption that millennial banking

Figure 3: Typical experiment card documenting a validation experiment about millennials' preferred communication channel with banks

ASSUMPTION	Millennials prefer to communicate via standard messenger apps rather than contacting a call center
EXPERIMENT	Pose a typical challenge to the test participants and offer them the option to contact the bank via WhatsApp (a proxy for a standard messenger app) or via call center to address it
TEST POPULATION	Millennials (initial test population size of 100 participants, add 50 participants per additional test round)
METRICS	• Measure which channel is preferred by the test participants • Measure which channel leads to getting the challenge addressed • Measure how many test participants switch from WhatsApp to call center after they fail to initially address the challenge
DECISION THRESHOLD	Accept the assumption (all conditions must be met) • 70% of the test participants chose as first communication channel WhatsApp • 80% of the test participants get the challenge addressed via WhatsApp Reject the assumption (any condition) • 70% of the test participants get their challenge solved by contacting the call center, either as a first or a second option • At least five test rounds have provided inconclusive results Inconclusive result • Perform an additional round of experiments

COSTS	RESOURCES	IMPACT
Low • Reward test population with small gift	Low • Agents answering challenge questions via WhatsApp and calls to the call center	High • Communication preference is determined • Satisfaction (challenges solved) is included in experiment

customers prefer using messaging apps, like WhatsApp or Facebook messenger, over communicating with their bank through an anonymous call center. The focus of the experiment is getting enough insights so as to be able to decide whether to offer messenger app-based or call center-based support to millennial customers.

3. PRACTICAL PERSPECTIVES

To illustrate key traits from using design thinking in a real-world setup, I consider the case study[3] of a medium-sized retail bank seeking to attract new customers. They wanted to extend their customer base with new entrants into the labor market, so-called millennials, receiving their first salary and no longer living with their parents. The challenge to address was formulated by the bank's board as introducing a new 100% mobile phone app-based banking offering for millennials. Any interaction with the bank should go through the to-be designed app. This challenge is a typical wicked problem ideally suited for design thinking. It focuses on a specific customer segment and its jobs-to-be-done: mobile banking. Although a lot has been written about millennials, it remains an open question how they, and not the bank, define mobile banking, especially with respect to need-to-have features. Applying design thinking in a purist way would have addressed the decision to go for a mobile phone app-based offering as part of the design process, based on observations and validated through experiments, rather than as a given.

3.1 Team

The first challenge the bank addressed, once the project got board approval, was to build a team. A review of available internal resources and capabilities concluded that significant external expertise was needed. They decided to assemble a core design team of around a dozen people, including strategy consultants, user experience designers, and software developers proficient in app development as well as back-office integration. The bank decided to staff the project office, including the overall project manager, with in-house employees. Very important to success, the board was tightly associated with the project. Coordination meetings with key board members and the design team were held on a weekly basis. The tight interaction with the board ensured critical buy-in at the most senior level of the organization. In addition to deciding on the next steps and guiding the project, the coordination meeting was authorized

[3] The presented case study is based on a real-world application of design thinking. It is presented in an anonymized way to be able to describe the highlights and challenges faced more candidly. The description is solely based on publicly available information. Neither the author, nor its employer, was involved in the described case.

to release fund and resources, if and when needed. Although this may worry some managers accustomed to thorough business cases and annual budgeting processes, design thinking embraces allocating funds and resources in a just-in-time way, notwithstanding a sizable and focused case for action. This agile way of handling budgets is key in design thinking due to the agile nature of the method.

3.2 Location

The bank decided to co-locate the core design team in a so-called war room,[4] allowing for optimal interaction. The war room was based at the headquarters of the bank, ensuring quick communication with major internal stakeholders. An alternative would have been locating it at a branch currently well frequented by millennials. The board prioritized closeness to internal stakeholders over closeness to customers. This decision is sound as the bank is not designing a solution focused on face-to-face interaction, but rather an app-based servicing model. In addition, all partners not part of the core design team, like software developers or user-experience designers, were required to have their offices within less than two hours of traveling time and speak the same language as the design team. The reasons for these decisions were ensure speedy and smooth interactions.

3.3 Method

As most design team members were new to design thinking, the team decided to use ethnographic interviews with target customers as the primary means of identifying insights, rather than relying on passive observations. After interviewing around 100 informants, the design team, during the learning step, came up with a list of jobs-to-be-done sought after by millennials. The offering must:

- include a current account to which the employer of the millennials can wire the salary,
- offer the ability to pre-allocate cash to different spending goals, like paying the rent or the electricity bill, and savings targets, like buying a new snowboard or going on a trip to Vietnam,
- allow for transferring funds to friends and share part of the funds with their partner or other millennials living in the same residential community,
- support retrieving physical cash (yes, millennials still want to have the option to withdraw cash, as some locations they frequent, like music festivals, only accept cash) and paying via their mobile phone,
- permit paying bills via wire transfers,
- and, very importantly, millennials are not willing to pay for getting their core banking jobs done, although they expressed a willingness to pay for additional, more sophisticated added-value services.

Most of these are standard banking jobs. Especially interesting is the need for specific budgeting and savings functionalities. This shows that millennials have a very structured approach to handling their cash assets. Interesting also is the lack of requirements for international functionalities, like handling SEPA wire transfers. The requirement for a free offering led the bank to strategize around freemium business models [Kumar (2014)] in the context of banking privacy constraints. But the decision was deferred to a later stage, focusing first and foremost on new customer growth. Although unusual in the banking world, it is quite common in internet business models to defer the profit formula question to after having achieved a certain customer base.

Based on the identified jobs-to-be-done, the design team developed an initial minimum viable offering specification. By iterating through four steps, at times on a weekly basis, the design team evolved the concept of a mobile banking app offering into a launched product.

1. Build and extend a prototype of the mobile banking app, adding specific functionalities one at a time, rather than focusing on delivering a fully functional app.

2. Make the mobile app prototype available to a selected group of target millennials for testing.

3. Collect feedback from the test users by conducting individual interviews, focusing on issues identified, ideas for improvement, and suggestions for prioritizing new features.

4. Adjust the minimum viable offering specification, incorporating the feedback received.

5. Iterate back to step 1 until the mobile banking app is considered good enough for launch by the test users.

[4] Sometimes war rooms are also called greenfield or lab rooms.

Positive feedback from the target millennial users was considered a necessary condition to adjust the app or add new features to the prototype. In addition, any changes were screened for consistency with the business strategy associated with a 100% app-based banking offering. A typical feature that was included based on user feedback was the possibility to associate a personal photo with a specific budget position; for example, associating a picture of a snowboard to the associated savings positions.

Due to the tight and iterative design schedule, various functionalities were included as building blocks from third parties as is, rather than customized for the prototype. This led to one key challenge not being fully addressed ahead of the launch. The credit card provider used, which is the one the bank uses for non-app customers, does not allow loading the credit card on the mobile phone and use Apple Pay or Samsung Pay. This led to the offering having to include a physical credit card, departing slightly from the 100%-app based offering goal.

The offering was launched after only ten months and exceeded, according to the bank's own estimates, its expectations. Rather than consider the offering being completed with the launch, the bank decided to continue iteratively improving the mobile banking app and adding additional features, using the same four design thinking steps, although at a slower pace than during the pre-launch period and expecting to do so for quite some time in the future.

4. REFLECTIONS

Design thinking has proven itself as a successful method for solving wicked business problems. Nevertheless, it is important to note that not all business problems require a full-fledged application of design thinking. For some problems, where the solution is clear, using design thinking is even counterproductive. Sometimes, only individual steps, or elements from those steps, are necessary, like, for example, validating a solution with real users, rather than assuming the project team knows best.

While reviewing numerous design thinking projects, I observed five key insights that need considering when applying design thinking to problem solving:

1. Applying design thinking to solve challenges found in banking requires a business strategy beforehand. No problem-solving method can overcome the lack of strategic directions. As the cat said in Lewis Carroll's Alice in Wonderland [Carroll (1865)], "If you don't know where you want to go, it doesn't matter in which direction you go." This may sound trivial, but many challenges observed in banks exist because of a non-existing, unclear, or poorly communicated strategy.

2. The composition of the design team and its location environment are as important as the design thinking method itself. Getting either wrong significantly increases the chances of failure. Without people willing to think out of the box and embrace change, no design thinking project will succeed.

3. Design thinking is about customers and their jobs-to-be-done. Unless target customers and their jobs-to-be-done are well understood, design thinking projects will fail. Taking a customer-centric approach is the only way to address customer pain points and their thought-after gains.

4. Design thinkers will get it wrong the first time! Failure must not discourage. It should support learning from mistakes and improving in subsequent iterations. Design thinkers get multiple chances to succeed, but only if they are willing to learn from and understand failures. Iteratively moving toward a sound solution is at the core of design thinking.

5. Without decision makers, usually executives or board members, involved, design thinking projects will most probably fail. Ideally key decision takers should be an integral part of the design team. If this is not possible, the design team should at least involve decision makers in experimenting during the validation step.

5. CONCLUSION

Whether it is fintech, new regulations, or increasing customer demands, banks need to rethink the way they address wicked challenges related to designing and launching value-adding products and services that meet current and future customer needs. Design thinking has emerged as a highly effective and customer-centric method for solving these types of business problems. It is based on observing customers in their natural environment, prototyping ideas, and validating them with real customers in an iterative way, working towards the best possible solution. It helps banks to target their innovation activities towards profitable creativity around customer needs and avoid being disrupted by incumbents.

REFERENCES

Arnheim, R., 1969, Visual thinking, University of California Press

Baer, M., K. T. Dirks, and J. A. Nickerson, 2013, "Microfoundations of strategic problem formulation," Strategic Management Journal 34:2, 197-214

Baumgartner, J., 2015, Anticonventional thinking: the creative alternative to brainstorming, Erps-Kwerps

Brown, T., 2009, Change by design, Harper Business

Carroll, L., 1865, Alice in Wonderland, Macmillan

Christensen, C. M., T. Hall, K. Dillon, and D. S. Duncan, 2016, "Know your customers' 'jobs to be done,' Harvard Business Review 94:9, 54–62

Cross, N., 1982, "Designerly ways of knowing," Design Studies 3:4, 221-227

Cross, N., 2011, Design thinking, Bloomsbury Academy

Kelley, T., 2001, The art of innovation, Harper Collins Business

Kristiansen, P., and R. Rasmussen, 2014, Building a better business using the LEGO® SERIOUS PLAY® method, John Wiley & Sons

Kumar, V., 2014, "Making 'freemium' work," Harvard Business Review 92:5, 27-29

Lawson, B., 1980, How designers think, Butterworth Architecture

Liedtka, J., T. Ogilvie, and R. Brozenske, 2014, The design for growth field book, Columbia University Press

Liedtka, J., R. Salzman, and D. Azer, 2017, Design thinking for the greater good, Columbia University Press

McKim, R. H., 1973, Experiences in visual thinking, Brooks-Cole Publishing

Osborn, A. F., 1963, Applied imagination: principles and procedures of creative problem-solving, Charles Scribner's Sons

Osterwalder, A., and Y. Pigneur, 2010, Business model generation, John Wiley & Sons

Porter, M. E., 1985, Competitive advantage, The Free Press

Ries, E., 2011, The lean startup, Portfolio Penguin

Simon, H. A., 1968, The science of the artificial, MIT Press

So, C., and J. Joo, 2017, "Does a persona improve creativity?" The Design Journal 20:4, 459-475

Spradley, J. P., 1979, The ethnographic interview, Waveland Press

Spradley, J. P., 1980, Participant observation, Waveland Press

van der Pijl, P., J. Lokitz, and L. K. Solomon, 2016, Design a better business, John Wiley & Sons

TRANSFORMATION

DIGITALLY-DRIVEN CHANGE IN THE INSURANCE INDUSTRY – DISRUPTION OR TRANSFORMATION?

JEFFREY R. BOHN | Head, Swiss Re Institute[1,2]

ABSTRACT

While technology continues to insinuate itself into all facets of financial services at an exponential rate of change, the insurance industry, in contrast, faces a slow-motion parade of promise, possibilities, prematurity, and pared-down expectations. Digitization, the birth of insurtech, machine intelligence, which incorporates multiple tools that explore and exploit data, and the collection and curation of (orders of magnitude) more structured and unstructured data are changing (and will continue to change) the industry in material ways; and not always in line with specific predictions but with developing general capabilities that are fairly predictable. An interesting case in point is the first wave of unsuccessful attempts made by technology firms to disrupt the insurance industry in recent years.

This article describes, from a large (re)insurer's perspective, the trends and challenges related to how technology and society's digitization are irrevocably changing risk markets and insurance. Based on the described trends, I will suggest one nuanced response to the question of whether insurance is being disrupted and/or transformed, while highlighting the trends in the insurance industry catalyzed by society's digitization.

1. INTRODUCTION

In the early 1990s, I sat in a graduate-level econometrics class learning about asymptotic theories related to estimators. I remember raising my hand and asking, "How many observations does one need to be close enough to infinity to rely on these asymptotic estimators?" The professor didn't skip a beat and answered, "40." He had done a fair amount of work in the area of empirical macroeconomics and 40 observations constituted a large dataset at that time (most indicators were only available annually with a few available quarterly.) Fast-forward almost a decade to when I worked as a credit risk quant and developed a corporate-bond pricing model. In that project, I had nearly one million observations. I thought that was a large dataset; however, this dataset was nothing compared to what has become known as "big data." Financial quants eventually gained access to many millions of observations and "big data" was born.

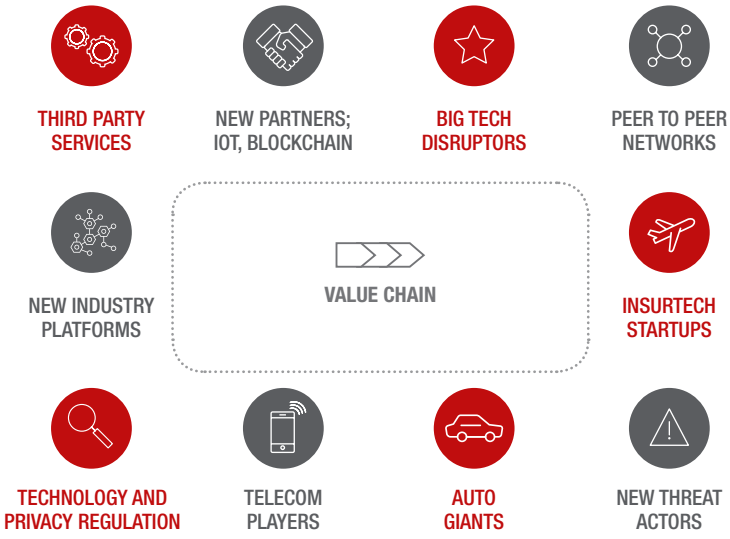
It turns out that "big data" eventually gave way to "data science" and today data-related work is often subsumed by the worlds of artificial intelligence and machine learning.[3] One popular technique in this arena is called deep learning. I am aware of the details of deep-learning

[1] Also affiliated researcher, U.C. Berkeley

[2] Many thanks to Jonathan Anchen, Darren Pain, and Binay Biswal for their extensive help on this article and to Leslie Sheppard for helpful comments. The views expressed in this article are those of the author and do not in any way represent those of Swiss Re or its affiliate companies.

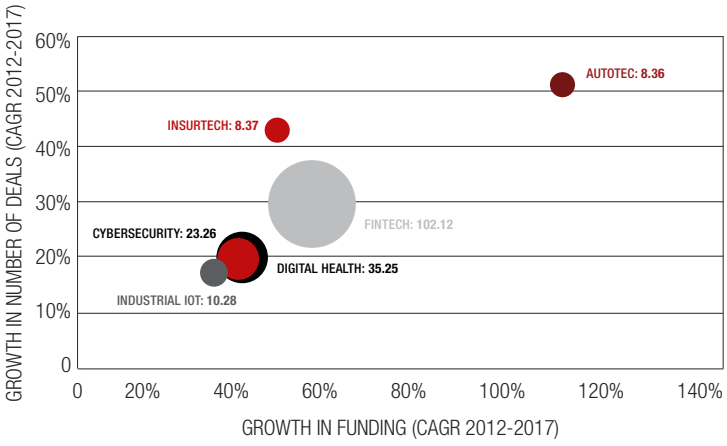
[3] As an aside, I first studied "statistics" in the early 1990s. This became "big data" and then "data science." As computational power increased exponentially, the same type of work came to be known as "machine learning." At some point, "artificial intelligence" supplanted "machine learning." As I will discuss later in this article, I prefer the term "machine intelligence." I heard recently that within some academic arenas, the field of ingesting, cleaning, analyzing, and visualizing data to support decision making is now called "statistics" once again. Unfortunately, the lack of clear and concise definitions of any of these terms continues to sow confusion. Later in this article, I will define these terms.

Figure 1: Impact of technology on the wider ecosystem



Source: Swiss Re Institute

Figure 2: Global tech start-up financing trends, 2012-2017



Sources: CB Insights, Swiss Re Institute

research projects using billions of observations – this is truly "big" data. An even more interesting twist to this increasing availability of data results from applying tools such as natural language processing (NLP) to unstructured data (e.g., text, audio, and video), which launches us into a world of unimaginably big data. (We should find another adjective to replace "big." Maybe "huge"?)

As the empirical research community races ahead to ingest and analyze quantities of data unthinkable 25 years ago, the parallel digital revolution that is transforming society in terms of how we shop, read news, investigate new topics, and find partners has been surprisingly slow in changing financial services in general and insurance in particular. This said, the past few years have been marked by a level of hype (not seen before) regarding technological change for banking, asset management, and insurance. Consultants, reporters, and CEOs throw around words like "AI,"[4] "blockchain," "platforms," "cloud computing," and "GPUs"[5] as if their institutions have already embraced a new "tech utopia." The reality on the ground is a mix of confusion as to what new technology tools to implement, constraints arising from legacy systems, capability deficits in terms of staffing, and concerns that digitization may have more costs than benefits.[6]

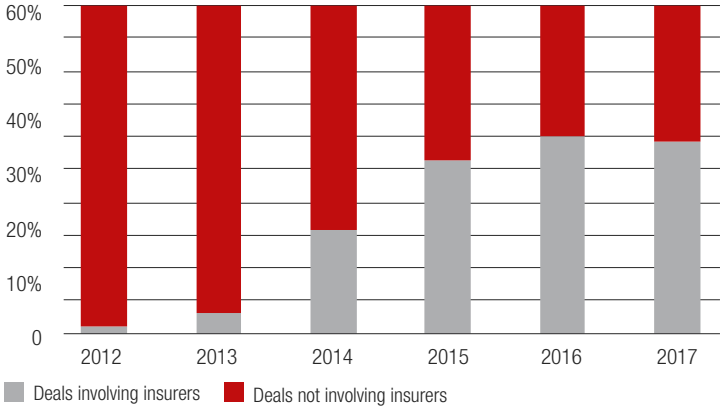
For readers who recall the tech bubble of the late 1990s, successful transformation of any industry almost never follows a straight, upwardly sloped line (in terms of improved efficiency, decreased cost, and increased productivity). The internet itself represents an excellent example of a technologically transformative platform that followed a complex path that ended up opening doors for large technology companies (aka Big Tech) sometimes known as the FAANGs (Facebook, Apple, Amazon, Netflix, and Google).[7] The windy path of how to make use of new tools like machine intelligence and productively exploit the growing pools of data will have dead ends, setbacks, and confusing forks. While this non-linear path causes stagnation and cynicism to build as a new world offers promise without productivity, the pay-off to patient and thoughtful implementers will most likely match or exceed the benefits we enjoy from similar technology-driven societal transformations (e.g., railroads, expressways, personal computers, telecommunications, internet, etc.)

With this background, I will describe in this paper how society's ongoing digitization in the context of the (re) insurance industry has produced an interesting mix of promises and pitfalls.

[4] Artificial intelligence
[5] Graphical processing units
[6] These observations arise from my personal experience across banking, asset management, and insurance. My comments in this context or in this article generally should not be construed as reflecting the views of Swiss Re or its subsidiaries. I am describing the typical circumstances found within financial services. While a few institutions will be outliers in terms of early and successful adoption of new technologies, the financial services industry as a whole has typically been slow to adopt new technology. I am drawing conclusions from many years as a leader of R&D units within large financial institutions and numerous advisory projects at the interface of technology and portfolio risk management, regulatory compliance, and business strategy development.
[7] Sometimes Microsoft is swapped for Netflix to produce FAMGA (Facebook, Apple, Microsoft, Google, and Amazon). Now that Google is part of Alphabet, it is harder to find a clever acronym that has three "As". I like FAANGs given its interesting contrast with the Chinese BATs discussed later.
[8] Lemonade Insurance Company is a property & casualty insurance company distributing insurance via online platforms without insurance brokers. They also claim to implement insights from behavioral economics using machine intelligence to improve efficiency in risk selection and claims processing (see www.lemonade.com).

Several years ago, the seeds of the insurance-industry disruption had seemed to be sprouting as new online insurance companies, such as Lemonade[8] (founded in New York City in 2015), appeared as data tools and machine intelligence were sweeping the world (or at least the technology world). Figure 1 shows that the impact of these tools seemed to extend beyond the insurance value chain itself to the whole business ecosystem in which (re)insurers operate. The combination of new business models, machine intelligence, piles of data, and a customer base ready for change seemed to imply an inevitable disruption to the (re)insurance industry.

Figure 3: Proportion of investments in insurtech startups with (re)insurer involvement



Source: CB Insight data on 600 insurtech deals

Today, the picture looks the same in terms of promise; however, the path to digital insurance utopia appears - and here is a spoiler with respect to the end of this paper and the primary conclusion - to have shifted from wholesale disruption via a bevy of insurtech startups to insurance value-chain transformation via an informed collection of industry incumbents. Insurtechs are most definitely part of this mosaic. Figure 2 shows that even though insurtech investment is still dwarfed by fintech investment, it is on a faster growth path. The barriers to entry for building insurance franchises have forced most insurtechs to shift toward the more reliably lucrative path to riches (in the insurance industry), which lies in facilitating transformation from within a large (re)insurer. This does not mean that would-be disruptors will not continue to fight the good fight; but rather, capital cushions, digital trust, brand awareness, and, most importantly, compliance with regulatory requirements create a daunting set of hurdles for a stand-alone startup to clear. Transformed incumbents have an overwhelming advantage. Figure 3 supports this conclusion as the increasing insurtech investments also reflect an increasing proportion of established (re)insurers' participation. The insurance ecosystem has added insurtechs as increasingly important members.

2. FRAMING THE DISCUSSION

One useful framework to determine whether digitally-driven changes to the insurance industry are disruptive or transformative focuses on the insurance value chain (see Figure 4).

Figure 4: Insurance value chain

PHYSICAL VALUE CHAIN	PRODUCT DESIGN/ DEVELOPMENT	PRICING/ UNDERWRITING	MARKETING	DISTRIBUTION	POLICY/CLAIMS MANAGEMENT
DIGITIZATION	• Robotics/Telematics/ Internet-of-things (IoT)/wearables offer usage-based insurance opportunities • Emerging risks such as cyber • Social-network insurance groups	• Use of Big Data/ analytics to identify new claims drivers • Predictive/prescriptive underwriting techniques • Artificial intelligence (AI) to hone risk assessment	• Position insurance as more customer-centric • Increase frequency of interaction • Use Big Data/analytics for micro market segmentation and personsalisation	• Customers prefer multi-touch, omni-channel interaction • Smart devices • Less face-to-face engagement • Scope for gains in efficiency in offline channels • AI-driven Robo-advisors	• Use of Big Data to reduce fraud and improve claims processes • Self-service apps to improve customer post-sales experience • Blockchain applications for smart contracts and claims administration
VIRTUAL VALUE CHAIN	INFORMATION CAPTURE AND ANALYSIS				

Source: Swiss Re Institute

[8] Lemonade Insurance Company is a property & casualty insurance company distributing insurance via online platforms without insurance brokers. They also claim to implement insights from behavioral economics using machine intelligence to improve efficiency in risk selection and claims processing (see www.lemonade.com).

To date, the primary focus of rapid development within the insurance industry and within insurtech startups is distribution. How customers buy insurance and manage their claims requires some kind of interface. In the successful digitization transformation of other industries, such as consumer retail, many useful and productive advancements in how customers interact with an interface is most straightforwardly transferred to insurance. Thus, most insurtech startups focus on this part of the value chain. The relative importance of attacking the entire value chain (and consumer experience), not just distribution, marks an important reason why insurtech startups have been forced to become part of transforming incumbents rather than radically re-defining the industry.

3. DATA, MACHINE INTELLIGENCE AND BLOCKCHAIN

Before drilling into the insurance industry's value-chain foundation to understand how digitization is producing transformation, a short detour into the current state of technology and data offers important insights into the changing digital landscape.

3.1 Data

Returning to the discussion of data: the capacity and tools available today to process structured (e.g., quantitative indicators) and unstructured (e.g., text documents, e-mail, etc.) data open new doors of opportunity. Data constitutes the new oil. As such, all industries, and especially insurance, should carefully consider the infrastructure, organization, and talent needed to monetize data. The comparison can be described as follows:

OIL	DATA
EXTRACT	COLLECT
REFINE	CURATE/COMPUTE
DISTRIBUTE	COMMUNICATE

Some efforts are underway within the insurance industry to collect data beyond what is already a data-intensive process to feed insurance pricing models and improve risk selection. This said, (re)insurers have an unimaginable amount of useful – but currently unexplored and unexploited - data scattered throughout their organizations and business networks. In particular, unstructured data, such as text and voice, constitute a potential treasure trove of insight-generating resources. Using the oil metaphor, these data lie dormant in the metaphorical ground (i.e., business ecosystem) until they can be collected, curated, computed (i.e., processed in the context of an algorithm), and communicated.

A disproportionate amount of investment in the data arena has focused on algorithm development, i.e., the compute step where the data are transformed into usable nuggets of insight. We now have an amazing array of data science tools. Unfortunately, not enough time and resources have been focused on the data collection and curation steps. It is as if most companies have developed a high performance auto, but ignore impurities in the fuel that regularly cause engine failure. In IT-speak, the extract, transform, load (ETL), plus cleaning process has been mostly neglected.

The regulatory constraints regarding the collection and use of data creates a significant barrier to entry for a company newly entering the insurance industry. While the information technology (IT) infrastructure necessary to exploit the panoply of data science flooding the technology ecosystem are mostly not in place at large (re)insurers, a tremendous amount of new investment is flowing in this direction. Big Tech, startups, and new data science teams at large (re)insurers are part of these efforts to build modern data platforms that can handle huge datasets that are both structured and unstructured.

3.2 Machine intelligence

A critical part of any plan for digitizing insurance includes new technologies; specifically those related to machine intelligence, which I will use as an umbrella term for the following overlapping terms often bandied about:

- **Artificial intelligence:** computer systems designed to undertake tasks that usually require human intelligence.
- **Artificial general intelligence:** computer systems that can perform any intellectual task that can be performed by a human (and maybe has even become self-aware).
- **Machine learning:** computer systems that can observe and learn tasks rather than being explicitly programmed.
- **Deep learning:** a type of machine learning where a system can learn (infer) from data without ex-ante assumptions as to models or underlying frameworks.
- **Meta-learning:** another type of machine learning where the algorithm focuses on the learning process itself, i.e., learning how to learn based on estimating meta-parameters.

- **Cognitive computing:** computer systems that simulate the processes in the human brain (can become one basis for machine learning.)
- **Augmented intelligence:** computer systems that assist humans in a set of intellectual tasks or tasks that support intellectual tasks.
- **Expert systems:** computer systems that use databases of knowledge to provide advice.
- **Robotic process automation:** tools and systems that can replicate or "roboticize" repetitive back-end processes currently performed by employees.

These technologies tend to be data intensive, though sometimes a machine intelligence can learn via simulation once it understands a set of semantic rules. However, in many cases the rules are not understood or cannot be written down in a comprehensive manner, making a simulation approach impossible to implement in most cases. Every algorithm within this space does not work for every use case. Many current failed implementations arise from mis-matching algorithms and use cases.

3.3 Hype cycle

The Gartner Hype Cycle (GHC) [see Panetta (2017) for details] provides an interesting lens into where we sit today in general (not just in insurance) with respect to machine intelligence.

First a quick explanation of this interesting framework for analyzing hype and expectations:

- **Innovation trigger:** a new technology is developed, and expectations begin rising.
- **Peak of inflated expectations:** as media, venture capital, consultants, academics, entrepreneurs, and practitioners dive into understanding and trying to deploy the new innovation, expectations quickly rise, and the hype reaches a climax. Often along the way, proofs-of-concept fail as the new technology does not live up to the hype.
- **Trough of disillusionment:** multiple failures, failed companies, and the media stories that follow puncture the bubble and the pendulum typically swings too far toward pessimism with respect to the new technology's promise.
- **Slope of enlightenment:** a few brave souls pick through the wreckage and find the diamonds of insight to figure out how to realize the technology's potential.
- **Plateau of productivity:** with more reasonable workplans and more patient stakeholders, the technology is molded into something useful that leads to productive implementations and sometimes substantively changes an industry.

In a recent Gartner publication [Panetta (2017)], most of the machine-intelligence-related categories they define are listed as having moved past the (hype) peak of inflated expectations and are now well on their way into the trough of disillusionment. Here are a few examples:

- Deep learning
- Machine learning
- Cognitive computing
- Cognitive expert advisors

The GHC-based analysis summarized above reflects perception with respect to emerging technologies across all industries. Within financial services in general and (re) insurance in particular, we are just starting to tip into the trough of disillusionment. In a recent panel session at an AI in Fintech conference (held in February, 2018) where I was a moderator, I asked a broad range of financial services executives with technology responsibilities where, in their respective companies, they had seen any useful and profitable implementation of machine intelligence. Only one executive raised a compelling example with a narrow segment of work his team does related to customer analytics. Since machine intelligence has mostly seen its recent successes in the social media space, this example makes sense. I expect to continue to see in the short term many successful machine-intelligence enabled implementations to support better customer interaction. Unfortunately, the promise of these new technologies across the entire insurance value chain has still not been realized. Interestingly, my experience, and the experience of a broad range of financial executives, suggest we are still waiting for the slope of enlightenment as the industry slides into the trough of disillusionment.

3.4 Finding value in "boring"

One useful consequence of this inevitable technology cycle relates to what is sometimes called the "boring" components of end-to-end data systems that support machine intelligence - namely, data collection, curation, and processing. One emerging area in this "boring" arena is robotic process automation (RPA). Related to RPA is augmented intelligence, where repetitive activities that support human-generated analyses (e.g., sifting through many documents, analyzing large datasets,

reviewing previous analyses, etc.) can be automated to augment insight discovery. For example, a simple application of natural language processing can help to machine-read thousands of pages of legal documents to direct an analyst to those parts of the documents that have substantive information. These technologies target pain points in a financial firm's back office and in the data preparation process in a financial firm's R&D team. Most data analysts spend unproductive time collecting, diagnosing, cleaning, and filtering data. This curation step can be materially improved with the right kind of machine intelligence.

"Despite its novelty, the core of distributed ledger technology has proven to be solid. It combines the best of distributed computing, consensus algorithms, cryptography and smart contracts."

3.5 Digitizing ecosystems

At the industry-wide level, the trajectory shift from disruption to transformation coincides with digitizing ecosystems themselves following this GHC framework. In the past few years, the (re)insurance industry has found itself sliding into the trough of disillusionment with respect to digitization as new technologies have not had as large an impact as expected. This does not mean digitally-driven change will not materially change the industry – it will – but the speed and path taken will not be so easy to predict or navigate. Consumer retail, transportation, and manufacturing have already been materially transformed due to digitization. (Re)insurance and financial services in general have digitally lagged behind and have not followed as quickly or as straightforwardly. An under-appreciated barrier to entry for new (re)insurance entrants lies in the tacit knowledge that is typically diffused throughout a large (re)insurer. This knowledge relates to building and maintaining the portfolio of insurance contracts, defining R&D agendas, building customized solutions, complying with regulations, and managing through tail-risk (i.e., low probability, high severity) events.

3.5.1 BLOCKCHAIN

Another technology family that has been heavily hyped without results (except in the context of cryptocurrencies) is blockchain. This technology also finds itself alongside the machine-intelligence technologies sliding into the trough of disillusionment [Panetta (2017)]. That said, the core technology, especially when expanded to include computerized transaction protocols (often called "smart contracts"), points to a decentralized technology future, which will most likely radically transform the insurance industry and commerce more generally.

While most readers will have heard about blockchain, I will cite a concise definition to make this discussion clearer: "A blockchain is essentially a distributed database of records, or public ledger of all transactions or digital events that have been executed and shared among participating parties. Each transaction in the public ledger is verified by consensus of a majority of the participants in the system. Once entered, information can never be erased" [Crosby, et al. (2016)].

Building on a blockchain, conditions can be embedded as part of the computerized protocol. This becomes a "smart contract," which enables a contractual agreement to make payments among participating entities/parties once pre-programmed conditions are met. The generalization of this technology makes it applicable to a wide range of transactions including those involving insurance.

B3i is an insurance industry initiative focused on blockchain. As one of the initiative's participants has pointed out:

"Despite its novelty, the core of distributed ledger technology has proven to be solid. It combines the best of distributed computing, consensus algorithms, cryptography, and smart contracts. At B3i we keep assessing different platforms, the likes of Hyperledger, Corda, Quorum, and are now quite comfortable that technology will do the job. The much bigger revelation was that, creating an industry-wide, enterprise-grade distributed ledger solution for insurance, requires rock solid inter-company collaboration. The more technology takes care of the basics, the more people and co-operatives count" [Meeusen (2018)].

In many ways, blockchain is more of an organizational and behavioral science challenge than a technological one. Like machine intelligence, the lack of working business models using blockchain in the insurance

space does not temper the enthusiasm of those who understand the promise of this technology in changing the way insurance is sold and managed.

An important point to remember specifically in the context of insurance is the huge amount of assets and lives that are not insured. This is called the insurance protection gap. Some estimates place this in the order of tens of trillions of dollars.[9] One barrier to providing insurance more broadly is the high cost of instantiating in a new or developing insurance ecosystem the current value chain to handle emerging risks, insure risks in developing countries, or insure smaller exposures (sometimes called microinsurance). Blockchain can dramatically reduce the cost of distributing insurance coverage and managing claims (particularly in contexts where existing insurance platforms are inchoate). Consider the possibility of shifting completely away from paper to a mobile-device enabled, digitized insurance product. The positive consequence of this technology will likely be much wider adoption of insurance and a material reduction of the insurance protection gap.

With the insurance value chain as the framework and an initial understanding of the current state of data, machine intelligence, and blockchain, we can assess where digitization is taking the insurance industry. We first consider its promise(s).

4. PROMISE(S)

Digitally-driven change to the (re)insurance industry offers some intriguing benefits that define the promise of this new digitized world. This promise focuses on the following areas:

- Dramatic reduction of cost to distribute & manage (re)insurance.
- Increased customization of insurance due to more data and better analytics.
- Networked connectivity across individuals and things (often referred to as the "internet of things" or IoT) leading to many new technology-led innovations in light of the data and better analytics.
- Improved data protection as (re)insurers invest in better data-related infrastructure to manage the increasing amounts of available data.[10]

As digitization continues to permeate our society, the actual insurance model will transform. If we place the insurance value chain in the context of the changing business ecosystem, digitization's promise encompasses an even larger range of possibilities (to be discussed in the next section) as the insurance business model radically transforms. In Figure 3, above, we see that large (re)insurers are responsible for an increasing proportion of the increasing total investment in insurtech startups. These trends imply a number of possibilities.

5. POSSIBILITIES

As the overall insurance premia have fallen in recent years, (re)insurers have looked to a variety of strategies to differentiate their products and services. Digitally-enabled changes constitute an important driver of new possibilities in the following areas:

- Reducing the insurance protection gap using data and technology.
- Customizing insurance products and services.
- Reducing the cost of marketing, distributing, and managing insurance.

Identifying and developing new products and services that leverage (re)insurers' know-how, platforms, and networks.

5.1 Reducing the insurance protection gap

The majority (70%) of existing economic losses to valuable assets in the world are not insured [Swiss Re (2015)]. In the event of large natural disasters (e.g., earthquakes in Christchurch, New Zealand, Tohoku, Japan, and L'Aquila, Italy) the difference in economic recovery is heavily influenced by the extent to which valuable assets are insured. In countries and markets where traditional insurance infrastructure supporting the insurance value chain is not available, digital platforms such as mobile apps with insurance contracts eventually instantiated in blockchain could dramatically reduce the cost of entering or expanding a market. Regulatory compliance and proper liability portfolio diversification (i.e., capital allocation optimization) are still important barriers to entering new markets; however, large incumbents will be much more willing to go after these new opportunities with a digitally-driven lower cost base.

[9] For details, please refer to the articles in the references section published by Swiss Re and the Swiss Re Institute.

[10] Note that (re)insurers that do not keep pace with data-related infrastructure investment will not only be at a disadvantage with respect to their competitors, they may also be inadvertently creating more vulnerabilities to cyber attacks.

Parametric insurance contracts, where claims are paid on the basis of an objectively measured indicator (e.g., typhoon insurance where payout depends on a typhoon's reported severity on an externally transparent measurement scale) become much more feasible with data platforms. Again, blockchain could be an essential part of a parametric insurance contract, further reducing the cost of managing insurance claims. The computerized transaction protocol embedded in distributed ledgers could create a relatively automated way to monitor relevant indicators and trigger claims payments per the conditions defined in the protocol.

New and still emerging risks suffer from difficulties in collecting, curating, and analyzing data relevant to insurability. These emerging risks typically face questions with respect to categorizing the risk, estimating relevant probability of downside events, estimating the severity of the downside event, and developing a claims process that has a reasonable cost profile. Expanding insurance product possibilities often lies in addressing these questions related to insurability. Examples of new and emerging risks include insuring natural assets (e.g., coral reefs, mangrove swamps, urban forests, and deep-sea ecosystems) and cyber-risk.

Many new risks arise out of society's increasing reliance on digital systems. While the risk of cyber attacks is likely at the top of the digitally-related priority risk list for most executives, operational vulnerabilities that arise from the increasingly complex mix of overlapping digital systems and networks that constitute the backbone of our digitizing society should not be ignored. I expect to see notions of algorithmic malpractice develop as software engineering becomes more integral to all aspects of human life. If core components of code in an enterprise-wide software system for a company, piece of critical infrastructure (e.g., airport, shipping port, power grid, etc.), or government is poorly devised and implemented, the downside risk could become increasingly material. Just as we hold medical doctors to a higher standard of conduct for their work, software engineers involved in system critical code development should be held to a higher standard of conduct. As standards and regulations develop regarding coding and software engineers, identification of vulnerabilities will likely lead to new insurance opportunities. This constitutes a new and emerging risk that will most likely have insurance implications.

New data platforms that process both structured and unstructured (e.g., text, audio, and video) data become resources for solving these insurability questions. Machine-intelligence-enabled tools can become essential to extracting the right mix of information to facilitate the development of new insurance contracts to cover new and emerging risks.

5.2 Customizing insurance

Internet of things (IoT) reinforces a trend to attach insurance to products as opposed to focusing on individuals in the context of how insurance is purchased. Furthermore, insurance can be customized by leveraging the data conveyed from a product using IoT. For example, auto insurance can be embedded into the purchase of the car. This contrasts with the currently common process where an individual buys an auto insurance policy separately from the purchase of the auto. With IoT, the insurance value chain focuses more on the item to be insured and less on the individual independent of his/her interaction with the insured item. For example, with auto telematics,[11] insurance pricing can be dynamically changed based on how the car is actually driven. The important possibility to highlight in this context arises from the shift of assessing an individual's risk (often with incomplete data with respect to a driver's risk profile) as a driver to assessing on a much more objective basis the risk of how an insured item is handled. The combination of machine-intelligence enabled algorithms to relate actual driver behavior captured via auto telematics with empirical estimates of probabilities of loss events and loss severities creates the possibility of building much better performing insurance portfolios.

New insurance products can also be developed as any IoT-enabled item can be tracked and insured in a similar way. For example, expensive equipment used on an infrequent basis (e.g., underwater camera equipment, mountain bikes, etc.) could be linked via sensors to parametric insurance contracts that only charge premia when the equipment is used. Mobile devices such as smartphones further reduce the cost, and increase the marketability, of these kinds of use-based insurance products. The opportunity in this trend arises from the combination of most aspects of the new digital economy: mobile devices, inexpensive sensors, connectivity, data

[11] Telematics deals with the long-distance transmission of computerized information. Most cars now transmit data real time on driver behavior. Even without this functionality built into the car, a mobile device can provide the same kind of tether to transmit information.

collection, machine intelligence, and some kind of digitally-enabled smart contracts. The result is a higher degree of possible customization without materially changing (probably even reducing) the underwriting and claims-processing costs.

5.3 Reducing costs

Much of the insurance value chain requires obtaining and processing data. Insurers use data to develop models, price risk, and decide whom to target in their marketing efforts. Then applicants enter data for evaluation and processing. Once a contract is finalized, an insurance company typically monitors the environment and the insured. When disaster strikes, claims require interaction between insurer and insured. Again, data are entered and processed. Along this path, regulatory compliance, data privacy protection, and legal review regularly add additional layers of processing complexity. All these processes are costly.

In order to reduce costs, paper can be replaced with bits stored in a computer network, and monitoring, reviewing, and evaluating can be replaced with machine-intelligence-enabled (or more likely, augmented-intelligence-enabled) algorithms. An important aspect of this process relates to the data-ingestion step where data should be filtered, formatted, pre-processed (e.g., missing data filled in, outliers evaluated for plausibility, and underlying data relationships evaluated for likelihood and plausibility), and reviewed. Much of the promise of machine intelligence relates to the nature of data available to an algorithm for "training." Better data ingestion will improve the viability and success of machine-intelligence enabled algorithms. Machine intelligence and better data infrastructures may require fixed costs up front, but will most likely reduce costs associated with operational failures downstream. Furthermore, these types of investments facilitate new data analyses that make it feasible to assess process efficiencies and potentially identify new opportunities to expand an existing revenue line or develop new businesses.

As discussed previously, many processes in the insurance value chain rely on expensive teams involved in repetitive and tedious data-related and reliability-ensuring processes. RPA offers the possibility of replacing any repetitive, data-intensive, and manual process with an automated system. The result is most often reduced cost and increased efficiency. Blockchain-instantiated computerized transaction protocols can substantially reduce the cost of offering insurance in segments and markets where existing insurance value-chain is inadequate or non-existent. In particular, developing countries without the network and infrastructure to sell and manage conventional insurance can suddenly facilitate rapid development of new insurance markets using mobile devices. This further reduces the global insurance protection gap.

5.4 Identifying and developing new products and services

As (re)insurers shift to a fully digital ecosystem, the process of identifying emerging risks and developing products to insure previously uninsurable risks will become much less costly. More importantly, new data, better-curated data, and new technology tools (e.g., machine intelligence) facilitate insuring emerging risks and previously ill-defined risks. A few examples illustrate these new opportunities.

Returning to the parametric insurance space (previously highlighted as a means to reducing the insurance protection gap), new, customized products become feasible. For example, business interruption insurance contracts tied to typhoon indicators or wind-speed measures could be further customized to cover particular days of the week or particular times of the day. That is, a payout would be triggered only if the typhoon hit on days and times when the business generates the bulk of its revenue. Other examples include flight-delay insurance and highly focused insurance. Imagine a product tied to a smart watch that doubles as a dive computer so that a customer only pays for coverage when the insurer's system detects the customer is actually in the water diving. Thus, the contract could cover risks to equipment and/or the health of the diver only when they are actually diving. The point here is data and technology create opportunities to personalize insurance and make pricing and coverage more dynamic.

6. PREMATURITY

While the possibilities portend an amazing new range of insurance products and a dramatic transformation of the insurance value chain, optimism has been premature. For example, auto telematics have been around for quite some time without influencing the way auto insurance is priced, distributed, or managed. This said, recent reports suggest it may just be a matter of time, as the data connectivity for cars is predicted to grow from 12.4

million in 2016 to 61 million in 2020 [Gartner (2016)]. At the very least, the optimism for changes in terms of months, or a few years, has been off target. This said, the optimism is probably on target over a longer time horizon (probably on the order of another five to ten years). In the case of auto insurance, the change to auto insurance will likely coincide with the shift to self-driving cars, which is an entirely difference insurance proposition; frequent, low-value accident claims will transform into infrequent, high-value accident claims. These intertwined trends in the auto insurance arena highlight confounding factors that make confident trend predictions premature. That is, society's digitization may introduce structural changes to the way we work and live, such that not only does the way we insure a particular risk type (e.g., auto, life, health, etc.) change, but the actual risk profile of that segment may have materially changed also. In this context, forecasting trends becomes much trickier.

The slow changes to auto insurance are still faster than what we see in other insurance segments. Selling and managing life and health insurance continues to be inefficient, overly bureaucratic, and needlessly complicated. A few companies have introduced discounts for using wearables that measure certain types of exercise. The gap is still large between what is technologically feasible and supported by objective evidence (in terms of how a particular intervention, such as measuring number of daily steps taken with a wearable, can be applied). Moreover, many insurance companies still exhibit excessive risk aversion in rolling out new approaches that may be demonstrably better based on pilots and other research. In this context, what we can say with confidence is that many more experiments are needed to figure out what technologies and/or behavioral changes actually improve an insurable risk. This type of research continues to be a high priority for the insurance industry.

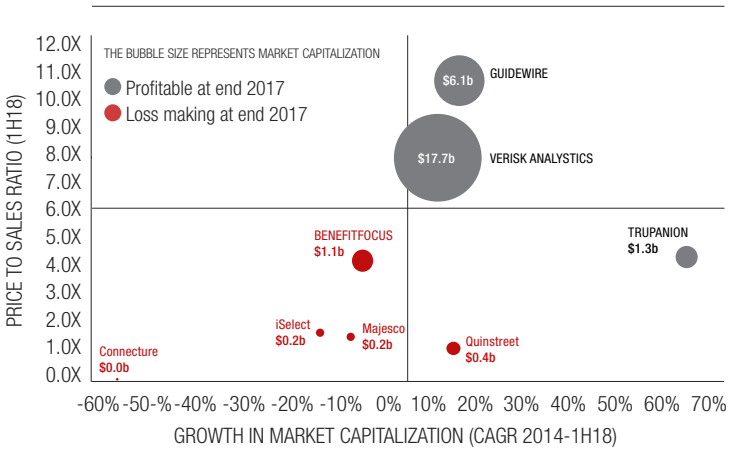
An interesting trend depicted in Figure 3, above, reflects an increasing number of startup deals undertaken in collaboration with a (re)insurer. In the early days of insurtech, the startup community appeared to follow a path focused on disrupting the industry in a way that displaced incumbent companies. They did not succeed. One likely explanation is that insurance industry regulatory constraints made incumbents investing in digital technologies the drivers of change. Whatever the cause, technology-induced transformation of the insurance industry now follows a path of digitally-aware incumbents working both with startups and internal teams to change the insurance value chain. Naturally, the speed of change slows when large companies lead the transformation. That said, partnerships and collaborations among large (re)insurers and technology firms appear to show the way forward. Insurtech's optimism was premature; however, the vision may have been correct when technologies are instantiated within incumbents with the proper regulatory compliance infrastructure, client network, market know-how, and strategic understanding. There have been very few IPOs in insurtech, suggesting that startups are keen to build long-lasting relationships with their investing insurers, conscious of the sector expertise the latter bring. For the few insurtech companies that have publicly listed, equity markets are demanding proof of a clear path to profitability. An analysis (see Figure 5) of firms that have listed in recent years shows that the few firms that are already profitable command better valuations. This leads to a more realistic assessment of today's opportunities, reflecting pared-down expectations.

7. PARED-DOWN EXPECTATIONS

A critical theme woven throughout this article relates to the promises and possibilities with the availability of a growing deluge of data; however, the systems, processes, algorithms, and know-how necessary to make this data deluge valuable are currently inadequate, making optimism premature. In particular, (re)insurance companies' IT infrastructures have not kept up with data availability. While most companies have constellations of projects focused on remedying this shortfall, the industry is mostly at the beginning of this path and still have far to go.

Expectations for customized insurance and the expectations for the plethora of promised machine-intelligence-enabled tools, products, and services have been pared down as the industry struggles with (and waits for) the completion of new technology platform developments. Fortunately, the advent of the cloud massively reduces the capital investment necessary to build new data platforms. Despite that, system architecture and proper tool choice become even more important to realize the productivity and product promise. Not all platform development teams are up to the difficult and complex task of designing, choosing components, and deploying a new enterprise-wide data platform. These new platform deployments now constitute a

Figure 5: Analysis of post-IPO performance of insurtech companies



Note: Growth in market capitalization for Majesco and iSelect is from 4Q15.
Source: Thomson Reuters, CB Insights, Swiss Re Institute

critical component of a strategy to navigate the digitally-driven changes over the next several years.

Digitally-enabled regulatory compliance presents a double-edged sword. On one side, RPA and machine intelligence (especially NLP) can lead to lower cost and more comprehensive regulatory compliance operations. Digitally-enabled regulatory compliance is often called Regtech. Insurance companies who invest in Regtech synchronized with regular dialogues with their respective regulators should see their costs decrease as their ability to regularly submit reports, respond rapidly to ad-hoc requests, and deal quickly with new requirements will materially improve.

On the other side, regulators will likely demand a higher degree of digital transmission and reporting. Many insurance regulators are already anticipating new reporting capabilities and developing proposals to expand the scope and depth of the data they require. This could lead to a more stable global system as regulators could potentially react much faster to systemically destabilizing circumstances that start with one or two institutions before these forces lead to a global crisis. This real-time reporting could also lead to difficulties for individual companies if errors creep into the transmission or regulators overreact to a particular day's set of reports. Companies that do not make sufficient investment, or make the wrong investment, in digital architectures may find their costs of regulatory compliance skyrocket as new requirements multiply.

The speed and investment in platforms throughout the industry are quite variegated. Despite the wide range of circumstances, these digital platform development projects share a few common characteristics: continual delays, failed enterprise-wide project implementations, and plenty of consultant-facilitated navel gazing. Reduced expectations appropriately track the continued inefficiency of insurance-industry IT. For example, Figure 6 shows that the lack of positive share price reaction for insurers that have been active acquirers of startups, relative to the less acquisitive insurers, implies the market remains to be convinced that insurers automatically derive significant competitive advantage by engaging with insurtech startups.

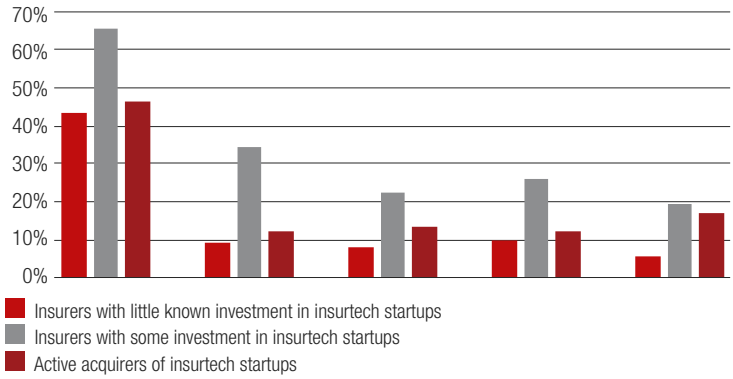
This problem plagues all financial services and is well recognized by most senior financial services executives. Eventually, these platforms will be built, and the data-driven projects will see a sudden injection of productivity and creativity from the new digital-ecosystem toolset. If history is any guide, the accumulation of years and years of gradual changes will suddenly reach a critical threshold and a sudden burst of non-linear improvement will lead to the realizing of these new technologies' promise. Despite the many false starts across the industry, leading incumbents are slowly moving forward. As I have explained above, I expect big and small technology firms will be a critical component of this incumbent-led industry transformation, also.

8. TENTATIVE CONCLUSIONS

In this article, I have described the recent past and current trends as digitization rolls through each aspect of the insurance value chain. Given how quickly environments and the digitally-enabled insurance industry ecosystem changes, I will finish with several tentative conclusions. Some trends have a long arc and will most likely continue on their current trajectories; other shorter-term trends could change quite quickly. Interaction effects and the rapid pace of innovation that deliver new tools and new business models on a nearly continuous basis creates drivers that can render a number of these conclusions obsolete in short order. This said, we can say something useful about these trend developments based on the information we have today. With these caveats in mind, I have identified four trends that illustrate this paper's tentative conclusions:

1. Leading (re)insurers have consolidated their position with respect to startups and now lead efforts to digitize the

Figure 6: Share price appreciation for insurers, classified by insurtech strategy, Jan-2013 to Aug 2018 (cumulative)



Note: Does not include insurers that are mutual or member owned (e.g., USAA, Mass Mutual).

Source: Thomson Reuters, Swiss Re Institute

insurance value chain, often partnering with or acquiring startups with useful technologies in the digital space. This trend shifts the leadership in adapting to digitization from disruptive startups to transformative incumbents.

2. Data have become paramount in any strategy to navigate digitization. While larger time series of structured data and efforts to find novel data continue to be an important component of this narrative, unstructured data (e.g., text, audio, and video) have become a new opportunity not fully exploited. Incumbents with proper tools and organization will differentiate themselves as novel data become a component of their competitive edge.

3. Machine intelligence and blockchain/smart-contracts are two over-hyped technology areas that have yet to be implemented in a materially profitable and useful way within the insurance value chain. That said, the collections of tools that fall into these two broad categories will continue to evolve. Eventually, a properly pruned subset of these technologies will trigger a discontinuous change in the way insurance is marketed, distributed, and managed. Much work remains in the realm of figuring out what specific tools in these spaces should be deployed. Both (re)insurers and their technology partners will benefit from deep investment and experimentation with machine intelligence and blockchain. The time horizon of successful deployment may be longer than originally promised, but the future will most likely include both of these technology families.

4. Regulatory compliance will continue to be a critical component of any strategy to leverage data and digital tools. This aspect of the insurance industry differentiates it from other industries in the sense that total disruption via a yet-to-be created startup is extremely unlikely. One area in this context that will be particularly onerous for any firm expanding its use of data, particularly in the area of personalization and customization – relates to data privacy. As an example of the challenges in the data area, the E.U. has rolled out a new set of rules for data privacy [called Global Data Privacy Regulations (GDPR)], in May of 2018. These types of new regulations will continue to come at a fast and furious pace furthering the advantage of large (re)insurers already equipped to manage compliance. Effective implementation of technology tools to facilitate compliance will be an important differentiator for forward-thinking (re)insurers.

Digital trust and brand reliability are intertwined with all four trends. As (re)insurers become the focus of clients at the time of processing claims, building a trusted brand becomes essential to maintaining and growing a sustainable insurance franchise. This "trust" will encompass how data are managed, how customer needs are met, the extent to which regulators are satisfied, and how a firm manages its capital. Digitally-informed brand strategies are another important investment for (re)insurers looking to navigate successfully the changing marketplace as technology permeates the insurance value chain. Deriving actionable insights from data with machine intelligence will be an important underlying differentiator for insurance companies struggling to adapt to the changes in the developing insurance digital ecosystem. Today, this transformation is led by informed-incumbents. While tomorrow could bring a new set of disrupters, the relentless onslaught of change will not go away: eventually every corner of the insurance industry's ecosystem will have to cope with society's digitization wave.

REFERENCES

Crosby, M., Nachiappan, P. Pattanayak, S. Verma, and V. Kalyanaraman, 2016, "Blockchain technology: beyond bitcoin," Applied Innovation Review 2: 6-19

Gartner, 2016, "Gartner says connected car production to grow rapidly over next five years," Gartner, September 29, https://gtnr.it/2MYD70D

Meeusen, P., 2018, "Distributed access to insurance," February 1, b3i.tech

Panetta, K., 2017, "Top trends in the Gartner hype cycle for emerging technologies, 2017," August 15, https://gtnr.it/1GichT4

Swiss Re, 2015, Underinsurance in property is a global, and growing, challenge, September 14, https://bit.ly/2FcmJS4

Swiss Re, 2012, "The mortality protection gap in the U.S.," September 5, https://bit.ly/2N6ohVB

Swiss Re, 2015, "Mortality protection gap report – Asia-Pacific 2015," August 25, https://bit.ly/2MEcBdA

Swiss Re Institute (SRI), 2018, "Closing the natural catastrophe protection gap," March, https://bit.ly/2NvMIZE

Swiss Re Institute, 2017, "Technology and insurance: themes and challenges," June, https://bit.ly/2NAcoED

THE CASE FOR A 21 MILLION BITCOIN CONSPIRACY

PEDER ØSTBYE | Special Adviser, Norges Bank[1]

ABSTRACT

Bitcoin and many other cryptocurrencies have currency-caps implemented in their protocols. Bitcoin is capped at approximately 21 million bitcoins. These protocols are complied by consenting operators. This paper discusses whether such currency-caps are illegal quantity-fixing conspiracies in violation of antitrust law. It is found that there is a present antitrust risk for cryptocurrency operators. This may render such operators subject to criminal and civil liabilities.

1. INTRODUCTION

Bitcoin and many other cryptocurrencies have currency-caps implemented in their protocols. Bitcoin is capped at approximately 21 million bitcoins. This protocol is complied with by the decentralized operators in the creation of consensual distributed ledgers. In an antitrust sense, this sounds like some sort of cooperation that may be subject to antitrust liability. This paper examines whether the 21 million cap implemented in the bitcoin protocol and similar caps in other cryptocurrencies are illegal quantity-fixing conspiracies in violation of antitrust law. It is found that there is a present antitrust risk for block-validators and other stakeholders involved in cryptocurrencies. This may render such stakeholders subject to criminal and civil liabilities.

Over the last few years, there has been an explosion of legal and regulatory research into cryptocurrencies and the associated technology more generally. This ranges from general assessments, as provided by Chuen (2015), Tu and Meridith (2015), and Paech (2017), to more specialized assessments, such as the legal status of so-called initial coin offerings (ICOs) provided by Zetzsche et al. (2018). Much of the literature gravitates towards financial regulation. This paper shares topics with Zetzsche et al. (2017), which assess the liability of participants in a distributed ledger. Zetzsche et al. (2017) rebuts the claim that the operators of distributed ledgers are outside the reach of the law and regulators. This paper also shares topics with Østbye (2017), which discusses competition policy for the cryptocurrency markets in general, also emphasizing the possible liability of the operators. In this paper we will, however, explore the narrow issue of whether the currency caps in cryptocurrencies are antitrust conspiracies. To the author's knowledge, this is not well explored in the literature. To make an adequate assessment of this issue, it is necessary to delve into the "nuts and bolts" of cryptocurrencies as provided by, inter alia, Narayanan et al. (2016).

[1] This paper should not be reported as representing the views of Norges Bank. The views expressed are those of the author and do not necessarily reflect those of Norges Bank.

2. CRYPTOCURRENCY TECHNOLOGY AND THE ROLE OF CURRENCY-CAPS

Bitcoin was launched in 2009, but documentation was already available in 2008. The creator or creators of bitcoin are unknown to the general public. The bitcoin white paper, Nakamoto (2008), was written under the pseudonym Satoshi Nakamoto. The intention behind bitcoin expressed in the white paper is that "[w]hat is needed is an electronic payment system based on cryptographic proof instead of trust, allowing any two willing parties to transact directly with each other without the need for a trusted third party." As a disruptive innovation and from the perspective of competition, it is a welcome potential challenger to banks and other financial service providers.

Many of the cryptocurrencies introduced in the aftermath of bitcoin seek to improve upon its shortcomings. For instance, scale and increased anonymity have been popular features to improve upon.[2] Some cryptocurrencies have been created by known natural or legal persons, and some even have mechanisms including more or less centralized governance and permission-based access. For instance, Ripple is intended to improve the efficiency of settlements between financial institutions.[3] Many new cryptocurrencies serve as utility-tokens to fuel service platforms. Ethereum is such a platform, providing a complete programming language on the platform, which can be used for, inter alia, smart contracts.

Cryptocurrencies are based on two main principles: cryptography-based asset disposal and distributed ledgers. Cryptography-based asset disposal means that cryptographic keys are used to sign transactions and verify ownership.[4] The transaction sender signs a transaction with a secret private key, and a corresponding public key can be used to validate that the transaction has been signed by the corresponding private key.[5] The cryptographic-asset disposal also allows for various mechanisms for conditional disposal, allowing for the execution of smart contracts. As it is private keys and not personal identities that determine control of assets, and there is no need to link real-world identities with private keys, the systems are pseudo-anonymous.[6]

However, digital assets are easy to copy, entailing a double-spending risk. A traditional solution is to rely on trusted third parties to maintain registers. The prime invention associated with cryptocurrencies is the elimination of the need for a trusted third party by letting the users validate transactions and maintain the integrity of the register. This is called distributed ledger technology (DLT). DLT protocols are designed to maximize the incentives of the users to maintain the integrity of the ledger in compliance with the protocol governing the cryptocurrency. The DLTs in various cryptocurrencies are designed such that they facilitate:

- **Detection:** the transparency of the ledger facilitates detection of dishonest behavior.
- **Punishment:** dishonest behavior is costly. The reward for validating transactions is given in the actual cryptocurrency, which will probably be lost in case of dishonest behavior. For many cryptocurrencies, the protocol allows for a reward for validation in terms of newly minted coins.

By such a design, users given the authority to validate transactions have incentives to do so honestly to maintain the value of the reward.[7]

The blockchain technology invented with bitcoin can be used to illustrate the implementation of such a design.[8] In bitcoin, put simply, each single transaction is broadcasted to the user-network and propagated according to peer-to-peer technology.[9] Participants in the system generate addresses from their public keys for transactions between them. The private key corresponding to each public key used to generate an address is needed to dispose of the bitcoins at that address. Competitive block validators collect transactions to add into a block to be added to the

[2] For instance, Litecoin seeks to improve scale and speed relative to Bitcoin. Dash, Cloakcoin, and Zcash, among others, seek to improve privacy. See Duffield and Diaz (2014) and Cloak (2018) for documentation of Dash and Cloakcoin, respectively. Both also improve scalability. Sasson et al. (2014) is the original whitepaper for Zcash. Improved anonymity is achieved in all three by coin-mixing arrangements that prevent transparency with respect to the sender and receiver of coins.

[3] See https://ripple.com/.

[4] Cryptography-based asset disposal is not an invention to be credited to cryptocurrencies. Public-key cryptography has been available for decades and has been suggested in variants of digital cash since the 1980s.

[5] The public key is generated from the private key with a non-invertible function, which is supposed to make this system secure. Non-invertibility is meant in a practical, not mathematical, sense. Advancements in technology may affect the security of the cryptographic functions applied today.

[6] However, as so-called network analysis can be used to infer identities from limited real-world information, several cryptocurrencies seek to improve anonymity by variants of mixing to hide the senders and receivers of transactions. See, for instance, Conti et al. (2017).

[7] This shares parallels with repeated prisoner's dilemma games, which are often utilized to analyze stability of cartels. See Belleflamme and Peitz (2015), Chapter 14.

[8] Although the description is aimed at being as precise as possible, some simplifications are necessary to avoid a too-lengthy description. For a more detailed description, see, for instance, Narayanan et al. (2016); for technical details, see Antonopoulos (2017). Alternative implementations of DLT, not based on blockchains, have also been developed as means to maintain the integrity of a distributed ledger. One alternative is to represent the ledger as a directed acyclic graph (DAG). IOTA is an example of a cryptocurrency using DAG for maintaining the distributed ledger as described in Popov (2017).

[9] Most software implements a rule that only valid transactions are propagated further to the network. However, this is not a hard rule, but dependent on users following the protocol.

blockchain. In bitcoin, the block size imposes limits on the number of transactions to be included.[10]

Each new block is pointing to a hash[11] of the previous block.[12] Hence, the blocks are chained together in a blockchain. A consequence of this is that if a validator wants to include transactions not consistent with the previous blocks in a new block, the validator would then need to alter the whole chain, back to a block consistent with the fraud, possibly the genesis block, to get hashes consistent with the present block of transactions. This could, in theory, be a simple task, but the bitcoin blockchain is designed such that this would be very costly. This will be explained next.

To be allowed to add a candidate block to the blockchain, the validator must be the first to solve a computationally costly puzzle. This puzzle consists of assembling the hash of the previous block, a hash of the transactions in the candidate block, some other inputs, and a freely chosen nonce into a hash-function, such that the resulting hash falls below a certain threshold. Hence, the validator must find a nonce that produces a valid hash consistent with the blockchain that the subsequent blocks will point back to. To solve this puzzle, the candidate block validator must perform many trials, as the hash function is not invertible and each trial contains minimal information about the solution. The lower the threshold, the harder it is to find a solution. To maintain the difficulty as the technological computational capacity increases, reductions in thresholds are implemented in the protocol.[13] The difficulty is set such that a new block is found on average every 10 minutes. The first finder of a valid nonce gets the privilege of adding its candidate block to the blockchain. However, it is not guaranteed to be a part of the blockchain. This depends on future block validators building their blocks on this particular block – that is, whether it becomes part of the consensus chain. Assuming that future block validators are honest and only build upon honest blocks, a validator has strong incentives to be honest and follow the protocol. Attempts to violate the protocol rules will render the block abandoned and the potential reward lost. This incentive scheme, based on the miners' use of computing-resources to validate blocks to receive a reward, is referred to as proof-of-work (PoW).[14] After the nonce is found, its validity is easy to verify, which facilitates the detection of dishonest behavior.

The incentive to be a block-validator is that the validator can include a fixed amount of newly minted bitcoins to a chosen address (normally of the validator itself or a mining pool in which the block validator participates) and transaction fees set at the discretion of the senders. According to the bitcoin-protocol, the reward of newly minted coins is halved at intervals of about four years.[15] This causes the total supply of bitcoins to converge from below at approximately 21 million.[16] The justification for this specific scheme is not provided by Nakamoto (2008).[17] The 21 million cap is not a technological limit; it is a consequence of the consented protocol followed by the validators. In theory, miners could be rewarded with newly minted coins forever, rendering the total supply non-capped. Actually, many cryptocurrencies do not have currency-caps, such as Ethereum[18] and Monero.[19] Some cryptocurrencies, such as the cryptocurrency Basis, aim to have stability reinforcing mechanisms built into the protocol to maintain a peg to another metric, such as the U.S. dollar. In such a case, the coin-supply will be floating to whatever is necessary to maintain the peg.[20] As the mining reward in terms of newly minted bitcoins declines, transaction fees are expected to increase in importance to encourage validation.[21] Since the block-validator is rewarded newly minted coins, the block-validators are commonly referred to as miners. The newly minted coin reward and the transaction fees are lost if the block does not become part of the consensus chain.

Cryptocurrencies are supposed to be decentralized. However, certain stakeholders may have more influential roles than others. As just explained, validators potentially

[10] The block-size is 1MB. A transaction contains on average 495 bytes, which makes the average number of transactions per block slightly below 2000.
[11] A hash function generates a non-invertible fixed-length output from an input in the same manner as a public key is generated from a private key
[12] To be precise: the header of the previous block.
[13] Although the main rule so far has been that the difficulty increases, it is also possible that the difficulty level reduces if the average time taken to find a new block increases.
[14] Various alternatives to PoW exist that may be used in combination with PoW. One commonly applied scheme is proof-of-stake (PoS). PoS means, simplified, that the block-validator is determined probabilistically according to the stake in the actual currency. For a detailed analysis of PoS schemes, see Bentov et al. (2016).
[15] The reward started at 50 bitcoins per block and halves every 210,000 blocks, which happens approximately every four years.
[16] Using a geometric series as an approximation, the upper limit is given by $21000 * 50 * \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k = 210000 * 50 * 2 = 21\text{m}$. Because bitcoins are not infinitely divisible, the maximum is slightly below 21 million.
[17] See https://bit.ly/1pQPBGe.
[18] To be more precise, this applies to the token Ether, see https://bit.ly/2zCJgVS.
[19] See https://bit.ly/2OfSVJB.
[20] See Al-Naji et al. (2018) for the Basis whitepaper. See also Østbye (2018b) for a critique of the mechanisms relied upon by Al-Naji et al. (2018).
[21] The limited capacity of bitcoin blocks can affect whether a miner includes the transaction in the block or not, or at least how fast the transaction will be processed. Huberman et al. (2017) studied equilibrium transaction fees in a simplified model. Tsabary and Eyal (2018) use simulations to show that validation only based on transaction fees can impede the security of bitcoin.

wield great influence. There is a risk of concentration among such validators, which would increase their influence.[22] If changes in the protocols are to be implemented, it is ultimately the validators that must execute these changes. Other influential stakeholders include so-called core developers. The formalized role of such core developers varies from cryptocurrency to cryptocurrency. Some protocols include mechanisms for awarding core-developers directly with newly created coins. In some permission-based schemes, the core-developers are fixed. The core-developers also have a role as the face of the cryptocurrency, resembling the administration of a corporation. Just as an administration might be replaced by a board, validators might replace the core-developers. Such influential stakeholders can be referred to as operators. However, due to the decentralized characteristics of cryptocurrencies, normal users holding a node may take part in the operation by propagating transactions and performing other functions, such as mixing coins to facilitate anonymity. The term normal users refers here to persons mainly using cryptocurrencies for the benefit they provide. The distinction between operators and normal users is not binary.

> "Cryptocurrencies are based on two main principles: cryptography-based asset disposal and distributed ledgers."

There are several ways users can acquire cryptocurrencies from their owners. Such acquirement can, inter alia, follow from bilateral private exchange, brokers, professional exchanges, and as payment for goods, services, and labor. In addition to those involved in the direct trade with cryptocurrencies, there is an ecosystem of third-party service providers, such as wallet providers for users to administer their cryptocurrencies, payment service providers, consulting services, and investment services. Such services allow for users not participating as nodes in the system, as such providers can appear as custodians for the users with their own nodes. Such custodians share similarities with banks and, in fact, some traditional banks are providing such services.

3. ARE CRYPTOCURRENCIES' CURRENCY-CAPS ANTITRUST CONSPIRACIES?

3.1 Antitrust conspiracies

The antitrust laws are legal rules regulating actions that restrict competition between businesses in the marketplace. Broadly speaking, the antitrust laws cover cooperation between businesses that restricts the competitive pressure among them, practices that might prevent competitors from competing fiercely in the marketplace, and mergers and acquisitions that restrict competition. Many jurisdictions follow the same template of competition law: prohibiting anti-competitive cooperation, prohibiting unilateral abuse of market power, and merger regulation that provides the legal basis for controlling mergers that restrict competition.

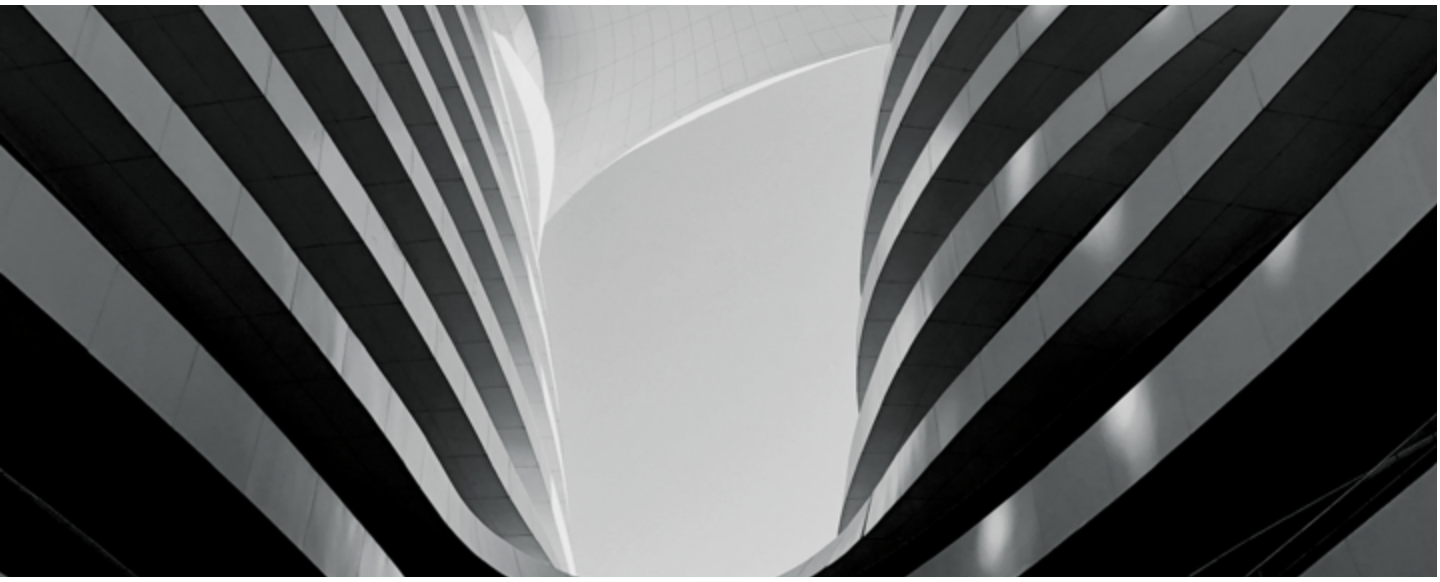
In this paper, we are concerned with the prohibition of anti-competitive cooperation, which also can be referred to as antitrust conspiracy. This paper will not delve into the details of any particular jurisdiction. However, the U.S. and the E.U. serve as examples. In the U.S., the Sherman Act, Section 1, prohibits "[e]very contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade." In the E.U., the TFEU Article 101 prohibits "agreements between undertakings, decisions by associations of undertakings and concerted practices which may affect trade between Member States and which have as their object or effect the prevention, restriction or distortion of competition within the internal market."[23]

No formal agreement is necessary to establish an illegal cooperation. However, some sort of "meeting of minds" is necessary to distinguish cooperation from unilateral behavior.[24] Cooperation can, inter alia, follow from some communication to facilitate the coordinated behavior. Individual rational adoption to the market

[22] A concern with decentralized validation is that validators or a coordinated group of validators could gain sufficient validation power to render a decentralized network de facto centralized. A so-called 51-percent attack refers to the situation where a dishonest validator or cartel of validators gains sufficient power to manipulate the ledger. A 51-percent attack is usually associated with so-called double-spending attacks. This involves a validator mining secret blocks to replace with the consensus blocks as the longest chain, facilitating the ability to spend the same coins twice. Much research has been devoted to the robustness of cryptocurrency protocols, in particular bitcoin, against attacks by validators with sufficient validation power. See, for instance, Conti et al. (2017) for a survey of possible attacks on the bitcoin blockchain. See also Narayanan et al. (2016), Chapter 5.

[23] For a more detailed description of competition law, see Østbye (2013).

[24] Unilateral conduct may also be subject to antitrust liability. This will not be discussed in this paper. For a general discussion on antitrust liability in the cryptocurrency markets, see Østbye (2017).



is not cooperation, even if the outcome is a mutually beneficial equilibrium among other "worse" equilibria in a game theoretical sense. For instance, two competitors maintaining an artificially high price (relative to cost) because both know that if one of them reduces their price the other will follow suit is not as such cooperation in an antitrust sense.

Cooperation that prima facie restricts competition may still escape illegality if it can be justified by legitimate grounds. For instance, in the U.S., cooperation not considered harmful per se, such as outright price fixing is judged according to a rule of reason standard, which means that it must be individually assessed as to whether the restraint is reasonable to make the society better off – that is, if consumer welfare is improved. In the E.U., the question is whether the restraint is necessary to realize social gains and the consumers receive a fair share of this gain. Another way to state legitimate grounds, which will be used in this paper, is whether the cooperation is ancillary to realizing gains that benefit society and consumers are not hurt. In this paper, we will consider consumers as those users using the cryptocurrencies for their intended purpose – that is, for transactions – without profiting from the operation as such.

Below, we will assess the conditions for currency-caps in cryptocurrencies to be considered as cooperations in an antitrust sense. The question as to whether such cooperation has legitimate grounds will be returned to in Section 4.

3.2 Are the operators liable entities?

The first question that must be addressed before we can take a stand on antitrust liability is whether operators of a cryptocurrency are liable entities according to antitrust law. In many antitrust regimes, such as that of the U.S., both natural and legal persons can be held liable. In some jurisdictions, like the E.U., only entities performing some economic activity can be held liable.[25] Such a restraint would, for instance, mean that those using cryptocurrencies only for private purposes, such as purchasing services for consumption, cannot be held liable. In such circumstances, the liability of operators such as block-validators, doing this as a hobby or for idealistic purposes, is unclear. However, for certain cryptocurrencies, such as bitcoin, many validators are clearly commercial, with business plans, employees, and investor backing. They will not escape antitrust liability on the grounds that they do not perform economic activity.

[25] However, at the national level, the member states may hold any person liable.

Cryptocurrency operators are normally scattered among jurisdictions. Hence, another question is whether the operators may escape liability because they are outside the jurisdiction of the countries wishing to apply their antitrust laws. Normally, antitrust liability is based on effects in the relevant jurisdiction and not the geographical location of the offenders. This means that if a conspiracy is established outside a country, but with effects in that country, possibly by persons with alien citizenship, this does not prevent the persons being held liable. This is rather an enforcement problem. As a country does not have enforcement powers outside its jurisdiction, the country is dependent on extradition agreements or if the person enters the country of jurisdiction voluntarily.

Consequently, it seems that restrictions on liable entities do not constitute any obstacle for antitrust liability. Lack of jurisdiction may, however, constitute a practical problem for enforcement.

3.3 Are the operators behaving unilaterally or in coordination?

For there to be an antitrust conspiracy, there must be coordination on the currency caps. Hence, the behavior of the operators cannot be a unilateral rational adoption. Despite common referrals to terms like “consensus protocol” and “consensus mechanism” in the cryptocurrency world, it is not obvious that the operators cannot be said to behave unilaterally. Rather, the governance structure of distributed ledgers is designed such that it is individually rational for each participant to follow the protocol without the need for communication or other coordination, as described in Section 2.

There are, however, several arguments that can be provided that indicate coordination. The original creators of a cryptocurrency may be a group of several persons. In this sense, there is coordination initially, and then new participants join this coordination. Furthermore, the protocol can be seen as an invitation by the original creators to participate, which is accepted by participants, thereby establishing coordination. Furthermore, as miners in PoW schemes join mining pools, each pool is a case of coordination. However, maybe the clearest indication of widespread coordinated behavior is the community communication between the operators in the operation of a cryptocurrencies. Operators communicate with each other for the coordination on protocol changes. This involves, inter alia, communications among core developers and validators, among validators, and among mining pools. This is done in community chat forums, and such communications are often a part of the protocol itself. For instance, in the bitcoin protocol, so-called bitcoin improvement proposals (BIPs) are a part of the protocol, and block-validators can use the blocks to signal their position [Narayanan et al. (2016), Chapter 7].

Consequently, it seems reasonable to conclude that there is coordination between the operators of a cryptocurrency in an antitrust sense, distinguishable from unilateral behavior.

3.4 Are the operators a company or structural joint venture?

A diametric opposite to unilateral conduct would be having the participants in a cryptocurrency be considered as a single entity, like a company. Normally, operations within a company will not be considered as illegal coordination. For instance, an owner of several shops may set common prices for all the shops without being subject antitrust liability. Furthermore, so-called structural joint ventures entered into by several parties operated as an individual unit with stable control-conditions, can be considered as a single unit not subject to antitrust liability for the operation of the unit.

A requirement of stable control conditions would render most cryptocurrencies outside the scope of being a unit under stable control conditions. Rather, the intention behind cryptocurrencies is that no one is supposed to be in control, although an oligopolistic structure of operators may prevent this intention in practice. Hence, most cryptocurrencies, such as bitcoin, cannot be considered a structural joint venture. This may, of course, change if a single validator or mining pool obtains sufficient computational power to de facto control bitcoin block-validation. If a mining-pool obtains de facto control over bitcoin on a stable basis, this mining pool may be considered as a structural joint venture. However, so far there is no evidence that this is the case.

As pointed out by Zetzsche et al. (2017), the conclusion might be different for permissioned special purpose cryptocurrencies, such as Ripple. Such cryptocurrencies often satisfy the condition of a stable control structure. This must be considered from cryptocurrency to cryptocurrency. For further discussion in this paper, it is assumed that we are not dealing with cryptocurrencies organized as structural join ventures.

3.5 Are cryptocurrencies' currency-caps restricting trade?

For a cryptocurrency currency-cap to be an antitrust conspiracy, it must restrict trade by somehow restricting the competitive process. It is well established that cooperation on quantity restrictions restricts competition. This is obvious in the "normal" economy of goods and services, as cooperation among suppliers to restrict output deprives the consumers of the benefit from suppliers competing with each other to capture market shares by, inter alia, lowering prices.

Cryptocurrency validators do not compete in terms of capturing market shares. In fact, users cannot choose their validators. Still, prima facie, it seems that coordination on a currency-cap restricting the amount of currency issued has the same effect. A cap on the currency increases its price in the same way as restricting output on normal goods and services increases prices. If competing validators could freely choose the reward for validating transactions, they may choose another reward than that set in the protocol, which would violate the cap. Hence, it is not unreasonable to assume that currency caps would be considered to restrict trade. This does not automatically mean that such caps are unlawful. This depends on the presence of legitimate reasons, as will be discussed in Section 4.

4. DO THE CRYPTOCURRENCIES' CURRENCY-CAPS HAVE LEGITIMATE JUSTIFICATIONS?

If we assume that cryptocurrencies' currency-caps compromise coordinations that restrict trade, the question is whether such currency caps can be legitimately justified. The exact legal assessment of such legitimate justifications varies between both contexts and jurisdictions. Such legal details are avoided here. In the present assessment, an agnostic approach is taken to the benefit of cryptocurrencies as such. If one takes the position that cryptocurrencies are bad for the society as such, no legitimate justifications may be found. Hence, the approach taken is that as long as there is demand for cryptocurrencies, they provide some sort of benefits to those who are involved with them. The question is whether currency caps are necessary to realize those benefits without harming the users, as described in Section 3.1.

4.1 Cryptocurrencies as money

Bitcoin and many other cryptocurrencies were introduced as alternative money and payment systems. Nakamoto (2008) makes several references to bitcoin as money and a payment system. At first sight it might appear obvious that the currency cap on cryptocurrencies is a prerequisite for their existence. Without any cap on the issuance, there is a chance that validators would issue too much, causing a value-loss and preventing the cryptocurrency from functioning as money – that is, from providing functions as mediums of exchange, units of account, and stores of value [Ali et al. (2014)]. Central bankers tend to argue that cryptocurrencies do not satisfy any of these properties today and, thus, are not money [Carney (2018) and Soderberg (2018)]. Such arguments may in some cases appear inconsistent, as it is at the same time argued that cryptocurrencies should be regulated for some of their money properties. Indeed, cryptocurrency exchanges are subject to money services regulations in several jurisdictions. Besides, central bankers' assessments of the moneyness of cryptocurrencies may not provide useful guidance on how they should be assessed under antitrust law. Furthermore, cryptocurrencies' capability to fulfill money functions may change in the future.

In the theory of private money supply, economists have argued that issuers' commitment to restricting issuance is essential for success [Klein (1974) and Fernández-Villaverde and Sanches (2016)]. Otherwise, the issuer would be tempted to issue too much, eventually causing the collapse of the value of the issued money. Numerous historical examples of privately issued money seem to confirm this thesis [Schnabel and Shin (2018)].

For cryptocurrencies, there would be an over-issuance risk with no restriction on validation rewards. For competing block-validators, there would be an externality present if they were free to mint whatever block-reward they wished, which would exacerbate the over-issuance risk. However, as block-validators need approval of their blocks by later block-validators to have their block included in the consensus chain, some discipline would be enforced. Later block-validators would probably be reluctant to include blocks with very high validation rewards. Such discipline would not be coordination as long as the block-validators make this decision unilaterally.

In developed economies, national fiat currencies are subject to inflation targets as well as constitutional checks and balances for money to remain credibly stable. Hyper-inflation seldom ends well in the countries where it happens. Hence, it seems that for cryptocurrencies to function as money, the issuance must be under some control.

However, controlling over-issuance is not equivalent to currency-caps. Inflation-targets for national fiat-currencies not only serve the purpose of protecting the currency from inflation, but also of protecting it from deflation. Deflation is not considered beneficial, as people may end up hoarding money instead of fueling the economy with consumption and investment. Consumption would be delayed, as holding the money would increase purchasing power. Investment would need to exceed the value-increase in money to be attractive. In macro-economic research, there have been various golden rules suggested for inflation targeting to protect a currency both from the evils of inflation and deflation [Langdana (2016), Chapter 11].

Consequently, it seems that if we are going to consider cryptocurrencies as money, some sort of money growth would be preferred to an absolute currency-cap to prevent harmful deflation. A concern would be that a coordinated rule on money growth would just be another coordinating antitrust violation. Such a concern has no merits, however, as a justified money growth rule more easily satisfy a legitimate justification requirement. The question then is whether a rule on money growth is achievable, or if currency caps are a technical necessity to restrict issuance. As several cryptocurrencies do not have currency caps, it seems they are not a necessity for cryptocurrencies.

As a currency cap seems neither optimal nor necessary for a cryptocurrency scheme, it seems plausible to conclude that the money character of cryptocurrencies is not a clear legitimate justification for a currency cap. This puts operators of such cryptocurrencies into antitrust risk. The short analysis provided here may of course be refuted by valid legitimate justifications, but according to standard burden-of-proof principles, it is the operators that must provide such justifications in an antitrust trial.

4.2 Cryptocurrencies as securities

Another way to look at cryptocurrencies is to apply the analogy to securities. Securities are typically bonds and equity stocks in companies. This analogy typically applies well to so-called initial coin offerings (ICOs) of tokens, where the tokens are a claim on a potential future value similar to securities. Many securities regulators have assessed whether securities regulation applies to ICOs [Zetzsche et al. (2018) and Fein (2018)]. There is no practice for considering caps on securities or company stocks as antitrust conspiracies. Such caps are usually essential for investors. Securities are claims on specific assets, such as a company. For instance, company stocks are residual claims on the value of a company. If new company stocks are issued, the value of the existing stocks is, according to theory, correspondingly diluted. A cap on the stocks, and the requirement of consent by the stock holders for diluting the stock by the issuance, is necessary for investors to acquire the stocks in the first place. Similarly, for bond issuers, if a debtor issues new bonds, the prospects of repayment in case of default reduces, as there are more creditors to share the remaining assets in case of bankruptcy. Hence, bond investors will normally require some control or commitments with respect to a debtor's issuance of new bonds.

Consequently, to the degree that cryptocurrencies are considered securities or share the characteristics of securities, there seem to be a weak case for considering currency-caps as antitrust conspiracies.

4.3 Cryptocurrencies as commodities

Cryptocurrencies could be considered commodities. The U.S. Commodity Futures Trading Commission (CFTC) has under certain circumstances considered cryptocurrencies as commodities [Adimi (2018)]. Comparing cryptocurrencies to digital gold is common, and is, in fact, used in Nakamoto (2008) to characterize bitcoin: "By convention, the first transaction in a block is a special transaction that starts a new coin owned by the creator of the block. This adds an incentive for nodes to support the network, and provides a way to initially distribute coins into circulation, since there is no central authority to issue them. The steady addition of a constant of amount of new coins is analogous to gold miners expending resources to add gold to circulation. In our case, it is CPU time and electricity that is expended. The incentive can also be funded with transaction fees. If the output value of a transaction is less than its input value, the difference is a transaction fee that is added to the incentive value of the block containing the transaction. Once a predetermined number of coins have entered circulation, the incentive can transition entirely to transaction fees and be completely inflation free."

Coordinations restricting the supply of commodities are at the core of what are considered antitrust conspiracies restricting trade. An international cartel restricting the supply of gold would be a strong antitrust case. The question is then whether bitcoin and other cryptocurrencies are so different from other commodities that a cap-coordination is justified. A particular feature of cryptocurrencies is that they are often pure digital goods not backed by any tangible assets. As opposed to gold, the scarcity of bitcoin and other cryptocurrencies is a pure social construct. Hence, they could potentially be supplied in an infinite amount. As discussed in Section 4.1, some scarcity is necessary for them to have value as money, which also applies if they are considered as commodities. Hence, it might seem that the operators would have some merits in arguing for a restricted supply where the cryptocurrencies are considered as commodities. However, this does not necessarily justify a currency cap.

This issue will not be concluded here. Rather, it will be asked instead what the theory of harm is. What theories of harm would the operators overcome in arguing for the legitimacy of the cap? The obvious theory of harm would be that the conspirators create something artificial in limited supply, and earn market power profits from any gains exceeding normal returns. There are, however, several problems associated with such an assessment. The first is the benchmark for normal returns. Some critical commentators would say that as cryptocurrencies have no fundamental value, all profits will be gains from the illegal coordination. However, this will be too-hasty a conclusion considering the possible benefits to society of cryptocurrencies. Another problem is that in PoW schemes, competing operators are not likely to obtain above-normal returns, as most of the income from mining is required to cover the costs associated with the PoW (such as electricity costs). The harm would then be the alternative cost to the society as the resources could have been spent better.

The quote from Nakamoto (2008) above also reveals another more sophisticated possible theory of harm. The newly minted coins work as an incentive scheme to validate transactions to later be replaced by transaction fees. This means that the users' transactions are subsidized in the beginning as validators finance their transactions by newly generated coins. However, as the cap is approached, and later reached, transaction fees must finance the operations. According to the theory of network effects and platform competition [Belleflamme and Peitz (2015)], this might be effective in obtaining sufficient scale. However, such mechanisms might also create lock-in effects, making it possible to exploit users in the future. A full analysis is beyond the scope of this paper [See Huberman et al. (2017) for a study of equilibrium fees in a bitcoin-like scheme].

Consequently, considering cryptocurrencies as commodities might initially provide a clear case of coordinated currency-caps as conspiracies violating antitrust law, which may render a burden of proof upon the operators to justify legitimate reasons. With that in mind, it also seems like the traditional theory of harm of antitrust conspiracies – excessive prices – does not follow the same mechanism in the operation of a cryptocurrencies. Competition among operators is likely to eliminate market power profits of operators. The theory of harm would be more justified if one thought of the entire cryptocurrency scheme as social waste, or if concentration of operators paved the way for excessive profits.

5. CONCLUDING REMARKS

This paper provides arguments and counter-arguments for currency-caps in bitcoin and other cryptocurrencies to be considered as antitrust conspiracies. If they are considered antitrust conspiracies, operators may be subject to both criminal and civil liabilities. Antitrust liability is probably just one of the many legal liabilities rendering it prudent for the creator(s) of bitcoin to remain anonymous. In the context of antitrust, Satoshi Nakamoto could be considered as the ultimate cartel ringleader. Legal liabilities are only theoretical without enforcement. For cryptocurrencies where the operators are pseudo-anonymously spread over a manifold of jurisdictions, enforcement is impractical. Hence, the real risk to the operators is probably marginal for now. This might change as operators become more concentrated and institutionalized, and as analytical tools improve in revealing the real identities of the operators.

REFERENCES

Adimi, A., 2018, "Regulating decentralized cryptocurrencies under payment services law: lessons from the European Union," Journal of Law, Technology, & the Internet 9:1, 1-1st.

Ali, R., J. Barrdear, R. Clews, and J. Southgate, 2014, "The economics of digital currencies," Bank of England Quarterly Bulletin 54:3, 276-286

Al-Naji, N., J. Chen, and L. Diao, 2018, Basis: a price-stable cryptocurrency with an algorithmic central bank," white paper, https://bit.ly/2NH00Wu

Antonopoulos, A. M., 2017, Mastering bitcoin: programming the open blockchain, O'Reilly Media, Inc.

Belleflamme, P., and M. Peitz, 2015, Industrial organization: markets and strategies, Cambridge University Press

Bentov, I., A. Gabizon, and A. Mizrahi, 2016, "Cryptocurrencies without proof of work," in International Conference on Financial Cryptography and Data Security, Springer, 142-157

Carney, M., 2018, "The future of money," speech to the inaugural Scottish Economics Conference, Edinburgh University, March 2, Bank of England

Chuen, D. L. K. (ed.), 2015, Handbook of digital currency: bitcoin, innovation, financial instruments, and big data, Academic Press

Cloakteam, 2018, "Enigma v2.1 A private, secure and untraceable transaction system for CloakCoin," white paper, https://bit.ly/2uJF5Xc

Conti, M., C. Lal, and S. Ruj, 2017, "A survey on security and privacy issues of bitcoin," arXiv preprint arXiv:1706.00916.

Duffield, E., and D. Diaz, 2014, "Dash: A privacy-centric crypto-currency," white paper, https://bit.ly/2p3gABX

Fein, M. L., 2018, "Bitcoin: how is it regulated?" white paper, https://bit.ly/2Mq9KQv

Fernández-Villaverde, J., and D. Sanches, 2016, "Can currency competition work?" working paper no. w22157, National Bureau of Economic Research

Huberman, G., J. D. Leshno, and C. C. Moallemi, 2017, "Monopoly without a monopolist: an economic analysis of the bitcoin payment system," working paper, https://bit.ly/2OhHLnA

Klein, B., 1974, "The competitive supply of money," Journal of Money, Credit and Banking 6:4, 423-453

Langdana, F. K., 2016, Macroeconomic policy: demystifying monetary and fiscal policy, Springer Texts in Business and Economics

Nakamoto, S., 2008, Bitcoin: a peer-to-peer electronic cash system, https://bit.ly/LjkXCv

Narayanan, A., J. Bonneau, E. Felten, A. Miller, and G. Goldfeder, 2016, Bitcoin and cryptocurrency technologies: a comprehensive introduction, Princeton University Press

Østbye, P., 2013, "Rational antitrust analysis: an inquiry into antitrust assessment principles and procedures," doctoral dissertation, Series of dissertations submitted to the Faculty of Law, University of Oslo no. 59

Østbye, P., 2017, "The adequacy of competition policy for cryptocurrency markets," working paper, https://bit.ly/2xaGgj0

Østbye, P., 2018a, "Will regulation change cryptocurrency protocols?" working paper, https://bit.ly/2p4OJBE

Østbye, P., 2018b, "Model risk in cryptocurrency governance reliability assessments," working paper, https://bit.ly/2N893QR

Paech, P., 2017, "The governance of blockchain financial networks," The Modern Law Review 80:6, 1073-1110

Popov, S., 2017, "The Tangle," white paper, https://bit.ly/2KXUlqQ

Sasson, E. B., A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza, 2014, "Zerocash: decentralized anonymous payments from bitcoin," in 2014 IEEE Symposium on Security and Privacy (SP), 459-474

Schnabel, I., and H. S. Shin, 2018, "Money and trust: lessons from the 1620s for money in the digital age," Bank for International Settlements No. 698

Soderberg, G., 2018, "Are bitcoin and other crypto-assets money?" Economic Commentaries No. 5, Sveriges Riksbank

Tsabary, I., and I. Eyal, 2018, "The gap game," working paper, https://bit.ly/2Oftd8a

Tu, K. V., and M. W. Meredith, 2015, "Rethinking virtual currency regulation in the bitcoin age," Washington Law Review 90, 271-347.

Zetzsche, D. A., R. P. Buckley, and D. W. Arner, 2017, "The distributed liability of distributed ledgers: legal risks of blockchain," University of Illinois Law Review, https://bit.ly/2OdnDT0

Zetzsche, D. A., R. P. Buckley, D. W. Arner, and L. Föhr, 2018, "The ICO gold rush: it's a scam, it's a bubble, it's a super challenge for regulators," University of Luxembourg Law Working Paper no. 11/2017

ARTIFICIAL INTELLIGENCE: CHANCES AND CHALLENGES IN QUANTITATIVE ASSET MANAGEMENT

FABIAN DORI | Quantitative Strategist, AQ Investment Ltd.
EGON RÜTSCHE | Quantitative Strategist, AQ Investment Ltd.
URS SCHUBIGER | Quantitative Strategist, AQ Investment Ltd.

ABSTRACT

Artificial intelligence has recently experienced a remarkable increase in attention, following staggering achievements in applications such as image, text and speech recognition, self-driving cars, or chess and Go tournaments. It is, therefore, not surprising that the financial services industry is working hard to improve investment decisions by incorporating self-learning algorithms into the investment process. However, do all sectors of the asset management industry exhibit characteristics that can benefit from applying artificial intelligence tools to uncover new patterns? Are there limits beyond which additional computing power and greater data availability have only marginal benefits? This article provides some initial answers to these questions. It demonstrates that the adaptivity and self-learning capability of machine learning tools could add value along the entire value chain of an asset manager. However, the inherently flexible nature of machine learning methods is also the biggest challenge. These methods must be applied thoughtfully and in the right context. The article provides a general overview of machine learning, elaborates on specific applications in quantitative asset management, and highlights limitations, challenges and possible remedies.

Artificial intelligence has in recent years received a lot of attention, following staggering achievements in various applications. It is, therefore, not surprising that the financial industry is also trying to improve investment decisions by incorporating self-learning algorithms into the investment process. Quantitative tools and algorithms have been used within the hedge fund industry to define systematic trading strategies for some time now, which is why quantitative hedge funds may provide a fertile ground for the application of new machine learning techniques. This article shows that the adaptivity and self-learning capability of machine learning tools may add value along the entire value chain of an asset manager. However, the inherently flexible nature of machine learning methods is also their biggest challenge. It requires that the methods are put in the right context and thoughtfully applied. This article begins with a general overview of machine learning, then elaborates on specific applications in quantitative asset management, also highlighting limitations, challenges, and possible remedies before concluding in summarizing remarks.

1. FROM MACHINE LEARNING IN GENERAL…

Machine learning refers to extracting knowledge from data by identifying correlated relationships without getting prior information about what causal dependencies to look for. It combines elements from both statistics and

computer science and has been in existence for many years. However, it has been mostly due to significant advancements in computing power and data availability that the application of artificial intelligence algorithms has become applicable for everyday life in recent years.

Most machine learning methods have been developed outside of finance and built on well-known statistical models, such as linear regression or clustering techniques. Still, the machine learning framework allows for much more flexibility. It can be applied to different kinds of problems, such as classification or regression analysis. Classification algorithms group observations into a finite number of categories, whereas regression analysis estimates outcomes of problems that have an infinite number of solutions. While machine learning is a very broad field, it can be classified into three main areas (Table 1).

The most successful field currently is **supervised learning**, where algorithms learn based on provided training data that reveal known relationships. The simplest form of a supervised learning algorithm is **linear regression**, which makes a prediction using a linear function of the input features. There is a general trade-off between optimizing the fit of a model on the in-sample training and the true out-of-sample prediction period. Given that all models tend to fit the training data better when more input variables are used, it may be reasonable to penalize additional model complexity in order to maintain sufficient generalization power for the prediction task. Methods such as "ridge" or "lasso regression" help in automatically detecting the most relevant input variables by regularizing model complexity to avoid overfitting. All three methods are by their nature linear, but may account for nonlinear relationships based on an appropriate manipulation of the input variables. A simple machine learning method that is not constrained to linear relations is the "k-Nearest Neighbors algorithm". This model looks for the k historical data points that come closest to the current situation and predicts future values based on these historical "neighbors". There are more complex nonlinear supervised learning algorithms, such as **decision trees** or **random forests**, however, they are not able to extrapolate and to make forecasts outside of the range of the training data.

Contrary to the methods described above, **unsupervised learning** algorithms only receive input data to learn from, but no information about the output data or relationships. These algorithms, therefore, detect patterns in the data by identifying clusters of observations that depend on similar characteristics. Machine learning can, for example, be used to identify the main topics in the news flow for a given stock. At the core of unsupervised learning algorithms is the idea of reducing dimensionality by clustering the data or by transforming it into simpler factor representations. **Clustering** methods partition the input data into subsets that exhibit common characteristics, such that the data points within a cluster share some notion of similarity that decisively separates them from the data points in other clusters. **Factor analysis**, on the other hand, relies on transforming the original data into the most relevant drivers or the most appropriate representation.

Table 1: Artificial intelligence and exemplary methods

SUPERVISED LEARNING	UNSUPERVISED LEARNING
Linear regression, Ridge, Lasso, k-Nearest Neighbors, decision trees	Clustering (k-means), Factor analysis (PCA, manifold learning)
REINFORCEMENT LEARNING	
DEEP LEARNING	
Multilayer, feed-forward neural networks	

Combining methods of supervised and unsupervised learning results in the so-called **reinforcement learning**, where the algorithm first detects patterns on its own, and then receives feedback from an exogenous source to validate or further guide the learning process. A reward feedback is required for the algorithm to learn a certain behavior.

The artificial intelligence literature also frequently refers to **deep learning** or **neural network** algorithms. This kind of method mimics, in a certain sense, the functions of the human brain by feeding information through different layers and nodes. The simplest form is called "multilayer perceptron" and can be seen as a generalization of linear models that perform multiple regression steps. There exist more advanced networks to deal with the challenges of simple networks and to allow for greater complexity. Given that even a sketchy synopsis of this area would be beyond the scope of this article, we refer the interested reader to the corresponding abundant literature.

Table 2: Artificial intelligence applications in quantitative finance

INVESTMENT PROCESS	POSSIBLE MACHINE LEARNING APPLICATIONS	EXEMPLARY METHODS
INVESTMENT UNIVERSE	• Identifying uncorrelated assets • Data mapping allowing for interpretation • Proxy assets with more liquid instruments	• Dendrogram • PCA, manifold learning • k-means
SIGNAL CREATION	• Create nonlinear forecast • Forecast based on predefined factors • Forecast without prior factor knowledge • Forecast based on new data sources	• Linear, lasso, ridge regression • k-nearest neighbors • Bag-of-words, term frequency-inverse document frequency
PORTFOLIO CONSTRUCTION	• Improve estimate of input parameters • Dynamically maximize target value • Scenario and stress testing	• PCA • Lasso, ridge regression • k-means
TRADING	• Pricing based on sparse data • Non-linear relations in transaction costs	• k-means • Lasso, ridge regression

2. …TO SPECIFIC APPLICATIONS IN QUANTITATIVE ASSET MANAGEMENT…

Each sector within the financial services industry uses artificial intelligence methods differently and according to their specific needs. In what areas of asset management can artificial intelligence provide added value? And what problems of investment managers may not be solved by such tools?

When we think about modeling the investment process – and hence about using machine learning algorithms to improve decision making – we like to subdivide the value chain into different steps. This allows for a systematic application of models that are appropriate for the specific task. We will provide details for the various steps in the next section, but start with an overview in order to facilitate the synopsis (Table 2). The first step consists of the "definition the investment universe". Next, the "alpha engine" or "signal engine" preprocesses the data in a proper way, calculates the signals for the various markets under scrutiny based on the models used, and maps these signals into the portfolio context. Subsequently, the "portfolio construction engine" or "risk management engine" builds the theoretical model portfolio based on a suitable algorithm, taking regulatory and investor specific limits into account. Finally, the "trading engine" translates changes in the model portfolio positions into effective trades.

What then, do these more general descriptions adhere to in more detail? Let us start with the investment universe, where machine learning tools may add value by identifying uncorrelated assets that provide true diversification benefits, or by mapping data into new representations that allow for other interpretations, such as the detection of style drifts in hedge fund strategies or factor exposures such as momentum or value. An appropriate tool for the first task would, for example, be a dendrogram analysis (see case study 1: Clustering the investment universe with dendrograms). The second goal may be achieved by relying on a principal component or manifold learning analysis. In a similar manner, artificial intelligence methods can be used to proxy the valuation or even actual investment of assets that have only sparse historical market data or are not eligible due to liquidity issues with more liquid instruments that appropriately mimic the characteristics of the actually desired assets. A useful variant to achieve that task would be the k-Nearest Neighbor model.

Case study 1: Clustering the investment universe with dendrograms

Dendrograms belong to methods of hierarchical clustering. The algorithm iteratively clusters first individual data points and then sub-clusters into a hierarchical order, depending on their respective correlation structure.

We use a dendrogram to structure a set of individual commodity markets into more meaningful clusters. Ideally, it comes up with the well-known sectors such as energy, precious and industrial methods for example.

At the bottom of the visual representation in Figure 1 are the single data points that are joined in first clusters. For example, the model groups copper and aluminium to a mutual cluster of industrial metals, or gold and platinum to precious metals. These two clusters are then joined to

form a cluster of metals in more general. Similarly, heating oil and crude oil are merged before being clustered together with gas oil as the energy complex. The energy and metals cluster are then put together to form a cluster of commodities that are more dependent on business cycle swings. The soft commodities soybeans, soybean meal, corn, and wheat are structured in a separate node that only consists of agricultural products. Interestingly, the natural gas commodity forms an individual cluster, most likely because of seasonality factors that separate it from the other energy commodities.

Figure 1: Dendrogram for investment universe

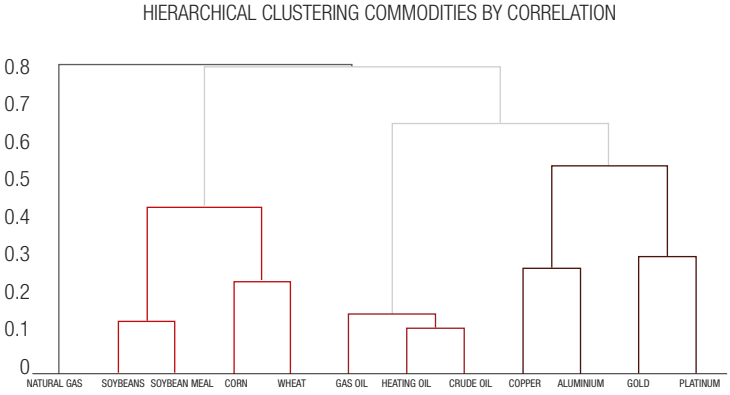


Figure 2: KNN for predictive power of moving averages

The aim of the alpha or signal engine is to produce forecasts on the direction and magnitude of future price movements of the relevant assets or their respective riskiness, and how to translate that information into a meaningful signal for the portfolio construction engine. Potential applications of machine learning methods in this field can be classified into three main blocks. First, an artificial intelligence algorithm may be helpful to create a nonlinear forecast based on a single time-series (see case study 2: Analyzing the behavior of the VIX index with KNN). Second, machine learning methods may derive the forecast value out of a predefined pool of relevant factors. More involved is the third application, consisting of letting the model select relevant input signals on its own or access new data sources in order to extract additional information.

Case study 2: Analyzing the behavior of the VIX index with KNN

The VIX index measures the market expectations for the volatility of the S&P 500 index over the coming month, based on index option prices. Given that volatility can neither become negative nor grow boundlessly, theory suggests a mean reverting behavior. Additionally, the distribution of changes in volatility levels is commonly skewed, mimicking the fact that spikes in volatility oftentimes occur very fast, while the slowdown in volatility normally takes more time and is a bumpier road.

Given that backdrop, we analyze the predictive power of two moving averages on past index movements for the future direction of the VIX Index based on a k-nearest neighbor classification algorithm. Figure 2 plots the values of the short moving average on the x-axis and the values of the long moving average on the y-axis. Conditional on the value of these two moving averages, the blue points represent moments in time where the future VIX index movement was negative, and the green points indicate future positive directional changes. While it is difficult to extract a meaningful interpretation from this scatter plot, a k-nearest neighbor analysis reveals further information. Based on this estimator, Figure 2 shows the decision boundaries for the two states of future directional movements in separate colors. The area colored in blue represents states where the two moving averages indicate falling VIX levels, whereas the red area stands for scenarios in which the two moving averages predict a rising VIX. Clearly, positive values for the moving averages are related to negative future VIX price movements, confirming the mean reverting

behavior after an increase in volatility levels. The picture for negative moving average values is more ambiguous, overall upholding the thesis of mean reversion, but also showing some signs of momentum. That makes intuitive sense, as volatility tends to trend lower after a sudden outburst.

> "Artificial intelligence aims to extract relevant knowledge from possibly unstructured data on a self-learning basis."

Based on these forecasts, the portfolio construction or risk management engine calculates the target positions, taking regulatory and internal restrictions into account. In this step, artificial intelligence methods may be helpful to improve the estimate for input variables. This could, for example, be achieved by reducing the dimensionality of the dataset based on clustering algorithms such as a principal component analysis. Instead of optimizing the portfolio with respect to a predefined objection function and specific constraints, machine learning tools may also be asked to tweak the portfolio in a more general way. For instance, by dynamically weighting the portfolio components such that risk-adjusted returns in the sense of Sharpe ratios or the ratio of average returns to maximum drawdown are maximized. Finally, enhanced scenario analysis tools may improve model validation and stress testing applications.

The trading engine finally translates the target positions into effective market orders. This step is especially relevant for large asset managers, as an estimated two-thirds of gains on trades are lost due to market impact costs when trading into and out of large position blocks.

3. ...TO CHALLENGES AND LIMITATIONS

In the previous section, we have highlighted various steps along the value chain of a well-structured investment process that in our view are suitable to be further enhanced by machine learning applications. However, there are a number of challenges and limitations that are not necessarily new to quantitative investment managers, but may be aggravated by the flexibility of new techniques.

Artificial intelligence aims to extract relevant knowledge from possibly unstructured data on a self-learning basis. It works especially well for tasks with precisely defined rules and stable probability distributions, such as mastering demanding games like chess or Go. However, the stochastic nature of financial markets with its lack of stable rules and probability distributions may challenge the validity of relationships that are learned from the past. Accordingly, models should always be applied to clearly defined problems and validated against sound theoretical assumptions.

Similarly, self-driving cars can be driven along the same roads as many times as is necessary to teach them all the relevant aspects and AlphaGo can play with itself until it perfectly masters the rules of the game. However, despite the seemingly abundant access to data, there is only one historical price trajectory for each financial market to train a model on. This limited data availability restricts the complexity of the artificial intelligence model that can be applied and, therefore, the flexibility of its output when forecasting future price movements. This problem is further exacerbated by the fact that the vast majority of data for financial markets has only been collected recently. As a consequence, researchers should focus on parsimonious model structures and not be misled by the mightiness of artificial intelligence models to adaptively learn the past.

Table 3: Challenges and possible remedies

CHALLENGES FOR ML IN ASSET MANAGEMENT	POSSIBLE REMEDIES
LACK OF STABLE RULES AND PROBABILITY DISTRIBUTIONS	• Apply models to clearly defined problems • Validate results against sound theoretical assumptions
LIMITED DATA AVAILABILITY	• Focus on parsimonious model structures • Account for mightiness of models to learn the past
LACK OF CAUSAL REASONING AND IMAGINATION	• Keep simplified transformation in mind • Diligent analysis before investing resources and modelling power
FUTURE REGULATION AND SUSCEPTIBILITY TO MANIPULATION	• Check future accessibility of data sources • Validate contributors to data sources

Next, machine learning models excel at identifying relationships in the data that may be unrecognizable to the human eye. Still, they lack causal reasoning and imagination that would be necessary to anticipate events that have not happened in the same way many times before. Would a trading model based on artificial intelligence have been able to predict the currency peg break between the Euro and the Swiss Franc by the Swiss National Bank in early 2015? Most likely not. In a similar manner, machine learning algorithms may just find theories that are already well-known and proven. While this confirmation may add comfort, it may also just be a waste of time and money. So, despite increasing computer power and data availability, it is necessary to keep in mind that quantitative models remain a simplified transformation of the world and will only have forecasting ability that is limited to specific tasks. Additionally, the complexity of calibrating artificial models requires a diligent analysis on where to allocate resources and model power most effectively.

Other more general potential limitations include future regulation and the susceptibility to manipulation.

4. CONCLUSION

This article provides a framework to assess the opportunities and challenges of applying artificial intelligence methods within a structured investment process. It highlights that the adaptivity and self-learning capability of machine learning tools may add value along the entire value chain of an asset manager. First, by more effectively using currently available data based on algorithms that learn to reveal new nonlinear relations or transform it into representations with more interpretable meanings. Second, by embracing new data sources that provide additional information. However, the inherently flexible nature of machine learning methods is also their biggest challenge. It requires that the methods are put in the right context and thoughtfully applied to solve questions that produce meaningful outcomes. It would be illusionary to believe that artificial intelligence will develop a profitable investment rationale on its own. Accordingly, we are convinced that machine learning will most likely turn out not to be the much-searched Holy Grail, but that it will help quantitative investment managers in further improving their allocation processes. Nevertheless, the application of artificial intelligence in asset management is still in its early days. This paper, consequently, provides evidence on first experiences, but no final results. We are looking forward to an exciting future.

NEW TECHNOLOGIES: DESTRUCTION OR NEW OPPORTUNITIES? OR BOTH...

THIERRY DERUNGS | Chief Digital Officer, Head Digital Solutions, IS Investment Solutions – Wealth Management, BNP Paribas sa

ABSTRACT

When people mention 'new technologies,' they are typically referring to the ones that transform, or create, business models, and have a deep impact on the organization and/or its processes. Other technologies could also be complex but are in fact just a day-to-day technical upgrade. New technologies are all about innovation together with disruption and cut across the entire organization. While the first key step is to start the process, the change itself needs to be managed more cautiously and steps must be taken to ensure it is methodical, coordinated, and instituted step-by-step. In this article, I will share our experiences in practice and walk through what we have done in the artificial intelligence domain.

I love to connect, speak, and exchange not only with my peers but also with many people. From CEOs to desk employees, from CTOs to project managers and developers.

I must say that if many people speak about new technologies opportunities, I observe quite often some fatalism or some wishful thinking. As I like to say, new technology is like teenage fantasies: many speak about it, few do it for real and often the first experience is disappointing. The good news is that you learn by doing.

I see CEOs who want to transform their companies but who do not go beyond a wish or a statement, and limit themselves to tangible short-term returns or efficiency gains while not considering new business models.

I see CTOs who have old legacies to maintain that are not compatible with the new technologies their businesses clamor for the next day.

I see CMOs who continue traditional marketing while trying some limited, and too isolated, web marketing activities.

I see project managers who have brilliant ideas but are blocked by unclear business cases.

I see developers frustrated to not be allowed to jump on new technologies.

Last but not least, I see employees afraid to be replaced by artificial intelligence or simply afraid of new technologies as they do not know what to do with 'that.'

You must think that I only walk on the dark side, and that I am one of those people who oppose new technologies.

No, I am a Chief Digital Officer living a success story in my company. However, I am also fully aware of all the issues, constraints, and challenges that a profound transformation raises.

For many companies, digital is fatalistic from the start. But they imagine that there must be some value somewhere as the others are moving forward.

Several CEOs told me that they must move to new technologies because their competitors are, or because they must reduce their costs. What a mistake.

This article is to deliver a message of hope. New technologies provide real opportunities and there are simple ways to apply them, without destroying or turning your company upside down. At least not from the start…

I wish to take you on this journey and demonstrate how you could also take your company, step by step, into the digital age.

Through simple steps, based on my 20 years of personal experience, I will walk you through the discovery of these new opportunities.

"We must go blockchain." Sure... and what color for your blockchain?

I love the cartoon in which a CEO is telling his CTO that they must build a blockchain. To check that he at least knows what he is talking about, the CTO asks him in which color he would like his blockchain. The CEO says that mauve has the most RAM.[1]

We all know about hype and the convincing power of consulting firms to persuade us that not using a given technology will soon put an end to our business.

If you are also a digital veteran, as I am, I am sure you remember the time when your company was not expected to survive without having a WAP (web application protocol) site. WAP was expected to replace html for mobiles towards the end of the 90s, but by 2001 its 2.0 version was already dead.

The technology itself was not the main issue. It was the fact that companies were creating useless and/or senseless sites on mobiles that was the real cause of failure.

Even if you are an IT firm, new technologies without a business purpose or perspective are just a waste of time and money.

You do not need to institute blockchain. You have new business models that demand new capacities, and while evaluating how to acquire them the blockchain technology could be one of the answers.

These few lines are to draw your attention to one of the key success factors. You do not need a new tech guru. You should avoid a dreaming business witchdoctor. You need a down-to-earth digital geek. You need that rare individual who can fluently speak in both languages of business and IT, who understands all the capabilities, and weaknesses, of the new technologies, and who can connect them to the new business models and use cases.

And to ease your hiring process, that person should also be a charismatic leader with extensive communication (and persuasion) capabilities.

That dual personality is the one who can highlight the potential business capabilities of the new technologies, opening a business use case dialog with IT.

They are also the best person to guide your IT through the technologies that need to be anticipated, and mastered, and how they should be best integrated.

1. YOUR FIRST CHALLENGE? JUST START...

The term 'new technologies,' typically in terms of digital, is often quite broad in context. Any check-up you do will reveal many opportunities for improvement (automation, analytics, client experience, etc.), new ideas, innovations, and unfulfilled expectations. Just pick just one.

As it is your first step, you must carefully choose THE use case. The one with a clever balance between simplicity and expectations.

For your selection, you should consider several factors.

- **Be simple and audacious:** select one simple, but visible, initiative that not only generates improvements for your clients but also, preferably, provides value to your employees as well. Ensure that you are in incubation, focusing on delivering an MVP (minimum viable product) rather than rushing for industrialization or turning your company's core business model upside down.

  And audacious? This is not always contradictory with simplicity. As your objective is to explore new capacities, your initiative must just be impossible, or very difficult, to achieve with your current capacities.

- **Be pragmatic:** you must be able to easily manage your initiative. No giant technical challenge, no extensive investments, no high risk to manage, no business revolution, just a small but bright idea to fulfill a need or fill a gap.

[1] https://bit.ly/2I7x199

Do not forget that you are a newbie. Be humble in your first ambitions.

- **Pilot:** you must be able to pilot your initiative within two to three months. It cannot be difficult as you have been simple and pragmatic. The pilot can simply be the delivery of the first part of your initiative, or a test with a limited target before a wider deployment. Whatever it is, its objectives are: test, learn, and adapt.
- **No (full) business case but a pragmatic and tangible business use case:** costs are pretty easy to evaluate, but the returns are not that easy to calculate. Do not stop yourself or slowdown because you cannot define proven and tangible returns for your first idea. As you have been simple and pragmatic, your investments are limited and so are your risks.

  Your business use case must solve a precise pain or provide, partially, a missing link.

- **Find the right partner(s) for incubation:** Partners for your ambitious use case must be wisely selected. You need a business matter expert who is (very) open-minded and fully cognizant of the fact that their usual way of working could be fundamentally changed by your initiative.

  An out-of-the-box thinker as the technical leader also plays a key role. No, your IT must not necessarily develop and/or master everything. The technical leader should support you to find the right partner for incubation, not for a full-scale industrialization. Your initiative is to allow your business to explore and learn new capabilities to deliver a use case. Working with a startup or a fintech is perfect for going fast, while acquiring knowledge, during incubation. Of course, you hope that this partner will also be the one for acceleration and industrialization. That is the second job of your tech leader: to already think about the industrial integration and deployment without slowing down your incubation.

2. WHILE STARTING, PREPARE YOURSELF FOR THE MESS

These two simple letters, AI (artificial intelligence), cover a very wide domain: NLP and NLG (natural language processing and generation), semantic, machine and deep learning, automation, big data, etc. The newness of 'new technology' is, of course, in the eye of the beholder: new for who? To be very clear, it is new for your company.

This is all about new capabilities for your company, which come with a huge challenge of integrating the underlying technologies with your legacy systems.

As it is worthless to have a car in your garage without using it in your day-to-day life (except if you are a very wealthy collector) any new capability must be integrated to enrich your current ecosystem. We all know how difficult it could be to combine (old) legacy systems with new technologies.

Difficult does not mean impossible but demands that you prepare your environment. From my point of view, there are two key work areas to invest in from the very first day.

Data is one of the most crucial for sure. Most new technologies are hungry for data, especially anything related to AI.

In addition, AI is a gourmet – it does not like junk food and requires gastronomic data. Data quality has always been important, but within AI it's even more so, since it takes you from the classic 'garbage in, garbage out' to 'garbage in, total mess out'. Data quality needs to be almost perfect for AI.

Although I am stating the obvious here, let me stress that data quality with a capital Q is your first key work area. Allow me to further explain and walk you through a simple example.

Many new technologies initially involve data crunching to build models, patterns, and/or knowledge. Beyond any new technology considerations, the simplest targeting work demands accurate data. Nevertheless, as a statistical analysis, data quality variation is smoothed inside the margins. Needless to say, that your target accuracy relies on decent data quality. This is already the case for any traditional business intelligence, hence no surprises here.

The key differences with the new technologies, especially while using AI in an analytic domain, are that you extend the number of data sources and of transformations (such as detecting intents in natural language to enrich your client data model), and that you are able to apply a model individually to all your clients, products, etc. Not only do the number of sources and transformations drastically amplify margin error, applying a pattern individually also diminishes the traditional mass tolerance.

Prepare yourself to discover that even if you manage your data very well, using new technologies will certainly create data quality issues. You need to be aware of your true data quality levels from the very first day and determine whether you could afford to improve them. These evaluations are crucial as they determine how far you can, or will be able to, dive into the new technologies.

It goes without saying that if your data quality is average, it will be a waste of time and money to start any business use case with new technologies.

And it will not do any harm to underline that this key work area is mostly a business duty.

So, you need quality data, what next? You have most probably heard many times that data is the new gold, but sitting on the richest gold mine is in vain if you cannot fully tap its potential and extract it at an affordable price. The opening of your mine is in my point of view the second key area to invest in as soon as possible.

This is all about API (application programming interface), hence I will not be discussing open banking or the Payment Services Directive (PSD2). My focus will be on the API-zation to open your legacy systems to new technology(ies) integration and easier data access. API is nothing new, but it becomes crucial as the standard way to connect and integrate any new technology.

Many executives view this transition as a basic and easy task for the IT. I should state that this is true and totally wrong.

Legacy systems are often accompanied by the good old MQSeries fellow (IBM messaging known as the software-glue), or other middle-layers. I must also state that Cobol is the IT Loch Nessie: many people think it has disappeared while in fact it is still widely present through legacy systems.

It is certainly the case that developing API on top of your systems is easy, though many in IT might respond by saying that it is not that obvious. However, while I agree that it does take time, the development itself is not rocket-science.

The real API challenges are not purely in the coding but in the strategy and the control you need. Indeed, new technologies are very often the open door to high volumes of data request that could lead to system overload and, potentially, crash. You need then to put a full API stack in place, usually referred to as an API Store, to set up all consumption controls and limitations according to your legacy capabilities.

If data quality was mainly a business work area, this one is much more for IT. Nevertheless, business must not be far away. Understanding business expectations in the short and long term is key, as usual, to set priorities for the API development and deployment.

3. AI'S ADVENTURES IN WONDERLAND

Pushing AI inside a company is not, always, an easy job. For sure you will be like Alice and (re)discover many Wonderland characters.

The Mad Hatter and the March Hare: they can discuss issues forever. Come back after some time and they will still be talking without acting.

The White Rabbit: for him, you should have delivered yesterday, or the day before. Fair enough, as time to market is always key, but what does he expect exactly?

The Queen of Hearts: too many consulting firms push you too fast to widen the projects, predicting the end of your business if you do not invest massively immediately, preferably with them. Indeed, you have to start, but you have to do so wisely and progressively.

The King of Hearts: he does not question what the Queen says. I hope for you that you do not have too many such characters in your company. We all know that, without denying that consultants can add value, they are there to sell their services.

The Dormouse: instead of questioning himself, he prefers to keep going as today for ever.

The Dodo: he is at the opposite of the Dormouse and wants to change through extreme actions without thinking about any side effect or impact.

Tweedle Dee and Tweedle Dum: I love them, since they always have a good story to tell to ensure that you will not open their Pandora box. They begrudge AI intrusion into their organization and processes.

The Cheshire Cat: he disappears each time you need a decision or have a question. He is there but not really.

The Caterpillar: find one. Always calm, thinking wisely, he works with you to move forward.

Please do not misunderstand me, I am not trying to scare you. However, you must be aware that AI, as with other new technologies, will change your company.

As your first challenge is to start, keeping an open mind about, and focusing on, all the potential difficulties and hiccups should guide you to make the right choice for your first move.

And the good news is that there are some safer AI domains to start with.

I will now provide some ideas and examples of practice. As these are based on our experiences, I will illustrate them with some of our partners. Be aware that we selected them based on our precise needs and expectations. These examples are not in any way product recommendations as there are no fit-for-all solutions.

4. ANALYTICS: THE NEWBIES' SAFE HARBOR

Analytics is easy: you just need a bunch of data and play with some tools. I am being a bit cartoonish even if it basically is as simple as that. As such, analytics is a safe area to start on AI.

Do not misunderstand me: safe does not mean restful.

In addition, we all need, and use, analytics to drive our businesses or operations. Starting with AI inside a well-known domain allows for easier discovery of new data approach added value.

Nevertheless, there is a compulsory condition before starting your AI analytic work. You must already have a

decent data environment, meaning data-warehouse(s) or data lake, with a good quality level. As mentioned previously, if you do not have such infrastructure or if your data quality is not at least good, do not start any initiative.

> "Let me stress that data quality with a capital Q is your first key work area."

Assuming that you have these, your IT should be able to build fast enough an isolated data environment which will be your playground.

As a first step, I would recommend that you select an existing model or client targeting. You then know which set of data you need (allowing your IT to prepare your first environment), you have an expert (the one who built the model or target and who knows the involved data), and you have a benchmark (your current model or target accuracy).

Finding the right partner to start with is then much easier. But, do not try to find the golden goose! Of course, you must search for a solution that will allow you to do more than just have your first experience, and hopefully one that will allow you to industrialize and integrate within your legacy. Nevertheless, this is a learning expedition and keep in mind that you could need another partner based on your newly acquired knowledge.

Our discovery of the artificial neural network analytics started exactly in that manner. Based on a set of existing target models and on some foreseen ones, we took advantage of our startup ecosystem and found DreamQuark.

Our first step was to do a fast (six weeks) first proof of concept (POC). Taking the most out of our agile and lean startup approach, our small team of business and data experts discovered and learned with DreamQuark what a neural approach could provide.

As the added value was promising, and as our partner seemed to be a perfect fit for us, we moved to a longer (four months) pilot. In that step, we decided to build a new model. But how do you benchmark it? As we had some sceptics, we launched in parallel a traditional statistical approach. It is important to highlight that it had also been the start of the in-depth technical assessment of the solution by our IT.

As the results came in, the model was built faster and much more accurately (in theory and in practice as we used the model with a set of our relationship managers). Last, but not least, our IT confirmed our capacity to integrate, at decent and affordable efforts, the solution in our framework.

Today, we are in the industrial deployment stage, not only in terms of the number of models but also across our different entities.

As you have seen, we have been able to move from a low risk and costs POC to the full acquisition of new analytic capabilities, based on a new technology and approach, in less than one year.

5. SECURITY AND FRAUD: AI GIFT OF X-RAY VISION

Security and fraud are unfortunately a common nightmare for all types of industry.

Regarding security, beyond the classic safety nets, such as firewalls, reverse proxies, etc., the key challenge remains to anticipate new attacks, such as identity theft.

Fraud detection typically relies on processes compliancy, complaints, and random, even if guided, controls. Again, the key challenge is not only to anticipate but also systematically control without having a controller per employee (and even then, who will control the controllers?).

For both, AI provides very high added value.

For example, I was worried by the growing importance of our corporate mobile devices in the day-to-day work of our employees. Of course, they were already secured, through a mobile device management solution, like other corporate applications. Nevertheless, my concerns were on the surroundings and the devices themselves.

Through one of my pilgrimages – I go regularly to events to meet startups and fintechs – I discovered Wandera and I was very interested in their security solution for mobile devices, as it seemed to fit my expectations for our employees.

Moving to a proof of concept (POC), then a pilot, demonstrated the added value of AI in identifying in real time patterns of risk, understanding behaviors, and reacting to deviations.

> "It goes without saying that if your data quality is average, it will be a waste of time and money to start any business use case with new technologies."

In this case, we left this highly specialized analytics to our partner, as it is based not only on our own traffic but also on its global footprint (the worldwide traffic from all devices connected to Wandera) and expert training on device, apps, and access points behaviors.

With regards to fraud detection, our approach is very different as we need to fully master internally the fraud patterns, since most of the requested data are highly sensitive and internal.

Our key objectives were not only to have the capacity to scrutinize every transaction but to also anticipate potential fraud cases. Again, these are all about patterns detection and behavioral analysis.

Our path and expectations were very clear as we started our journey with homemade developments. We were aware from early in the process of our limitations and started to search for a specialized partner matching our specific needs.

In this case, we moved to a mature startup, NetGuardians, as they matched a number of our key requirements: strong experience in finance, an on-premises solution, a model-based approach that we can fully master and control, and a behavioral- and pattern-based approach through AI and machine learning. As you might have noticed, knowing exactly the way to go, having pragmatic use cases, and possessing a vision are crucial for selecting the right partner or deciding on whether or not you should develop in-house.

In addition, it was also important for us to be able to deploy progressively not only across our different locations but also in terms of fraud detection scope.

Aligned with our specific expectations, NetGuardians' AI allows us to control all transactions with different perspectives and to combine them.

6. BOTS: THE CLASSIC

Bots are perhaps classic, however moving from a good old script bot, often a bit dumb, to a clever bot is not an easy path. Indeed, behind the 'bot' word are hidden several complex technologies: NLP (natural language processing), semantic and/or intent detection, machine or deep learning and, finally, sometimes, NLG (natural language generation).

The good news is that there are plenty of out-of-the-box solutions.

The bad news is that they too often promise magic when the bot training takes a lot of effort and time

In this case, we decided to dedicate our annual hackathon in Singapore to find a partner. Impressed by GoodAtlas, the winner, we took them to our Factory (our internal incubator and accelerator) to deliver not a classic bot but a voice bot dedicated to support our front people.

On our way to deliver a running MVP (minimum viable product), some five months into the project, we realized that they would not be able to support us in our industrial ambitions.

I want to underline that it had been a truly win-win situation. We had the opportunity to deep dive in all these topics and technologies and learned a lot, allowing us to find our new industrial partner much easier. For GoodAtlas, it had been the opportunity to learn too and, finally, to rotate their business model and focus on their key added-value.

And now? We are moving forward on industrialization with two different tracks.

The first one is to develop a homemade clever chatbot dedicated to our clients by using the API from a new external partner.

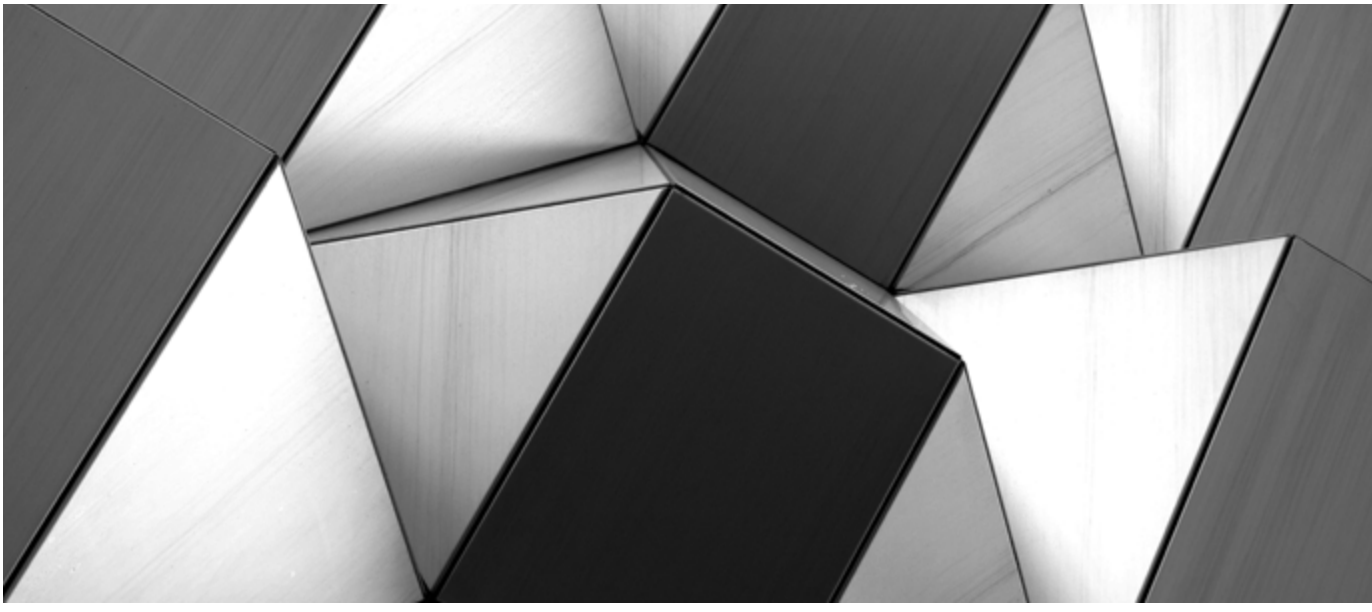
The second took over the work done in our Factory and develops with a fintech our voice bot to augment our relationship managers.

7. MY TWO CENTS…

I hope that sharing my experience will help you build yours.

My intention was to highlight the challenges you could encounter while starting on the new technologies, especially on AI.

Having gone through this experience, I can assure you that there are many ways to move from the Wonderland to the promised one, and to embrace the new opportunities of our modern, fast evolving, world.



THOUGHTS ON THE ECONOMICS OF BITCOIN

ERIK NORLAND | Senior Economist, CME Group
BLU PUTNAM | Chief Economist, CME Group[1]

ABSTRACT

Several basic economic concepts can be applied directly to understanding bitcoin. Bitcoin supply is highly inelastic. Many commodities, such as natural gas, just as an example, find it hard to increase or decrease production in the short-term when prices move quickly. Economists talk about this lack of price sensitivity in the short-run as an inelastic supply. And, inelastic supply is typically associated with substantial price volatility. In addition, taking an application from commodities, say the mining of copper, the marginal costs of production can be very critical to price dynamics. In the bitcoin world, the term used is "difficulty," due to the "mining" of bitcoin being based on math problems. The price feedback loops involve "difficulty" as a major driver of price, and price also influences "difficulty". Finally, we note that transaction volume may influence price trends, and rising transaction costs are a risk indicator for bitcoin.

1. INTRODUCTION

What is most striking about the economics of bitcoin is the juxtaposition of the certainty of supply and the uncertainty of demand. The rate at which bitcoin is mined has been highly predictable and, unlike almost any other asset, currency, or commodity, its ultimate supply is a known quantity, fixed in advance. There will never be more than 21 million coins. This feature makes bitcoin supply almost perfectly inelastic. No matter how high the price rises, miners will not ultimately produce any more than the prescribed amount. Moreover, price rises will not even necessarily incentivize a more rapid mining of bitcoin. Even if they did, it would mean miners create more bitcoin today at the expense of creating less of it in the future, since the total supply will reach a hard, asymptotic limit of 21 million coins, expected to be reached by 2140 or so, based on the mining algorithms.

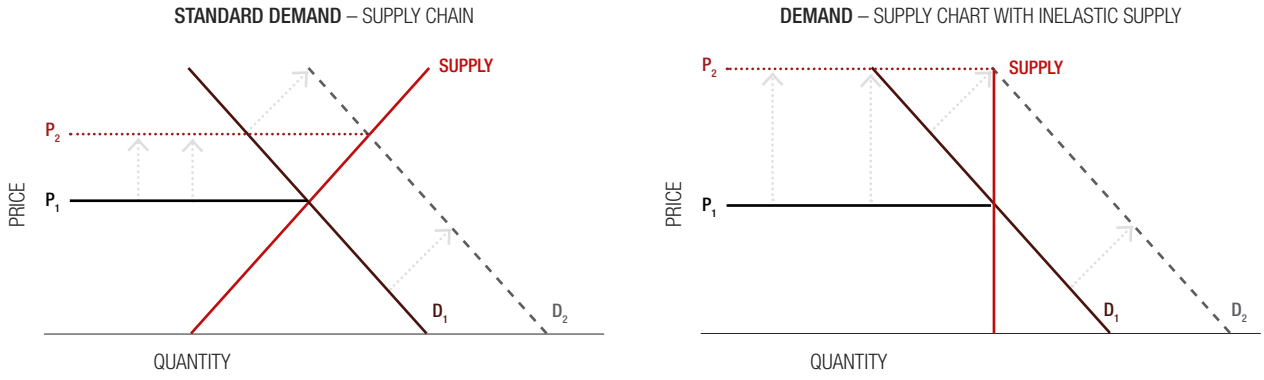
We analyze the economics of the bitcoin marketplace by finding parallels in the world of commodities to understand what it means to have an inelastic supply. Then, we move to the relatively more difficult task of demand analysis to complete the bitcoin economics picture.

2. ECONOMICS OF SUPPLY INELASTICITY

The supply inelasticity explains in large part why bitcoin is so volatile. Items with inelastic supply show a greater response to demand shifts than items with elastic supply. The same is true of demand: the more inelastic the demand, the greater the price changes in response to small fluctuations in either supply or demand. In the abstract example below, we show the relatively modest price response to an upward shift in demand for a market

[1] **Disclaimer:** All examples are hypothetical interpretations of situations and are used for explanation purposes only. The views expressed here reflect solely those of the authors and not necessarily those of their employer, CME Group, or its affiliated institutions. The information herein should not be considered investment advice or the results of actual market experience.

Figure 1: Elastic supply (left) is less price volatile than inelastic supply markets (right)



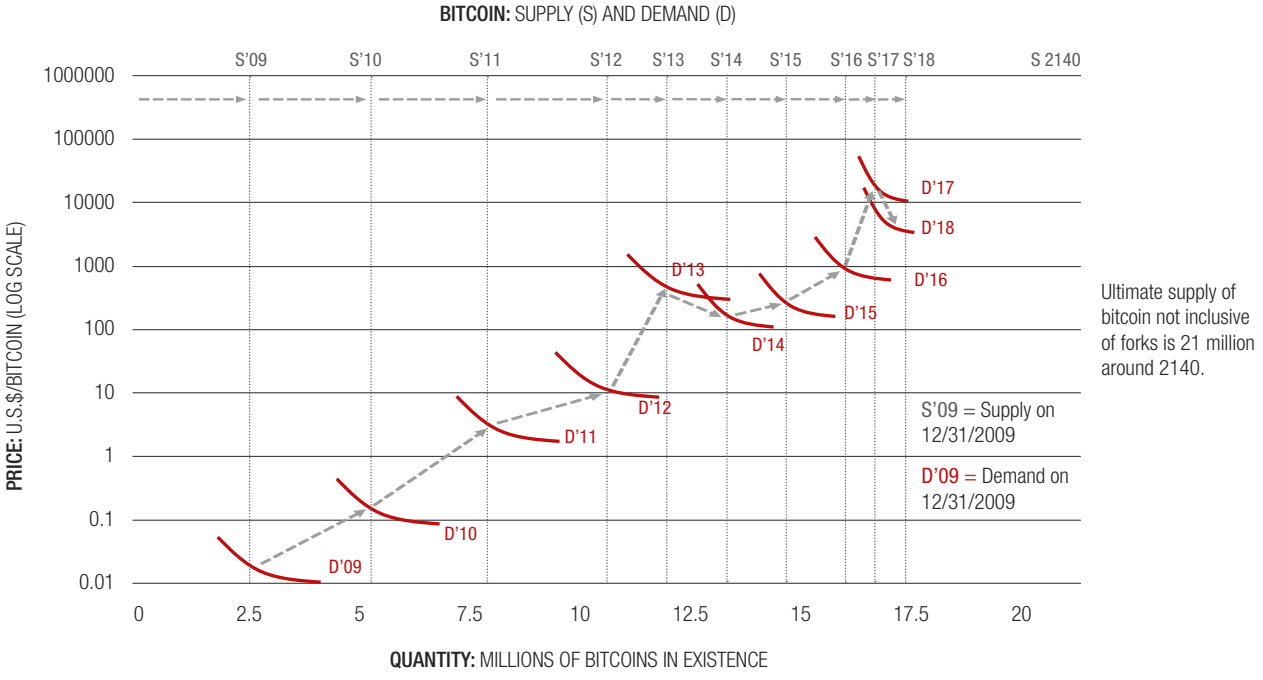
Source: CME Group Economic Research

with flexible supply elasticity on the left and contrast it with the much bigger price response from the same demand shift in a constrained supply market on the right.

Take as an illustration the case of natural gas. Natural gas is a classic example of a market with highly inelastic supply and demand. If prices soar today, consumers will still need natural gas to generate electricity, heating, and to fuel industrial processes; and they will be willing to pay for it, at least in the short term. Natural gas demand is, therefore, highly inelastic.

The same is true of natural gas supply. If prices double, which for natural gas is not all that uncharacteristic, producers will likely not be able to supply a great deal more of it in the short term. Similar relationships hold for crude oil, although are less dramatic. What differentiates the analysis of commodities like natural gas and crude oil

Figure 2: Inelastic expansion and slowing growth of bitcoin supply



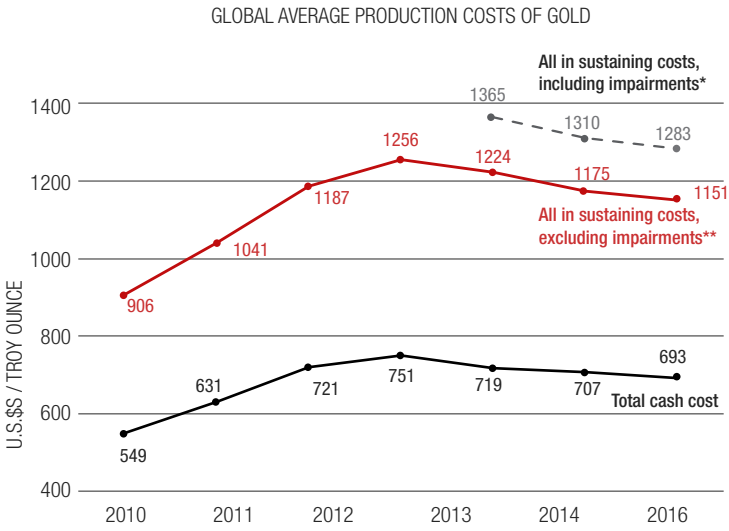
Ultimate supply of bitcoin not inclusive of forks is 21 million around 2140.

Sources: Bloomberg Professional (XBT), Blockchain.Info (supply), CME Economic Research calculations

from bitcoin is that their long-term supply and demand shows a meaningful degree of elasticity, even if the short-term supply is more about inventory swings than production adjustments. If natural gas or crude oil prices experience a sustained rise, producers can and will find ways of producing more of them – or at least they have so far in history. Meanwhile, consumers will find ways to use them more efficiently in response to higher prices. This is not the case for bitcoin directly, although rising prices might increase the probability of "forks" that split bitcoin into the original and a spinout currency such as Bitcoin Cash (August 1, 2017), Bitcoin Gold (October 24, 2017), and Bitcoin Private (February 28, 2018).

Bitcoin's limited and highly inelastic supply is also a major factor driving its price appreciation, a rise so spectacular that it can only be appreciated when seen on a log scale. In bitcoin's first four years, supply grew by roughly 2.5 million coins per year. Even then prices were rising as the user community grew. Since then supply has continued to grow but the pace has slowed substantially while demand has occasionally dipped, even on a year-on-year basis.

Figure 3: The cost of mining gold



Sources: GFMS Gold Survey 2016 & 2017, Metals Focus: Gold Focus 2015, Metals Focus Gold Mine Cost Service

Notes: *Retrenchment costs/carrying value write-downs; ** Cash costs + corporate administration (head office) + interest + exploration + sustaining capex

Bitcoin's limited supply and soaring price make it difficult to be used as a medium of exchange outside of the crypto currency space. Imagine one's regret if one uses bitcoin to purchase a mundane item such as a cup of coffee only to find that the bitcoin spent would have been worth millions of dollars a few years later. As such, investors treat bitcoin as a highly unreliable store of value – a bit like gold on steroids.

One often asked question is: will bitcoin replace fiat currencies such as the U.S. dollar? We think that the answer is a resounding no. Bitcoin's price is too unstable to compete as a store of value; bitcoin's transaction costs are too high and too variable for it to be used as a medium of exchange.

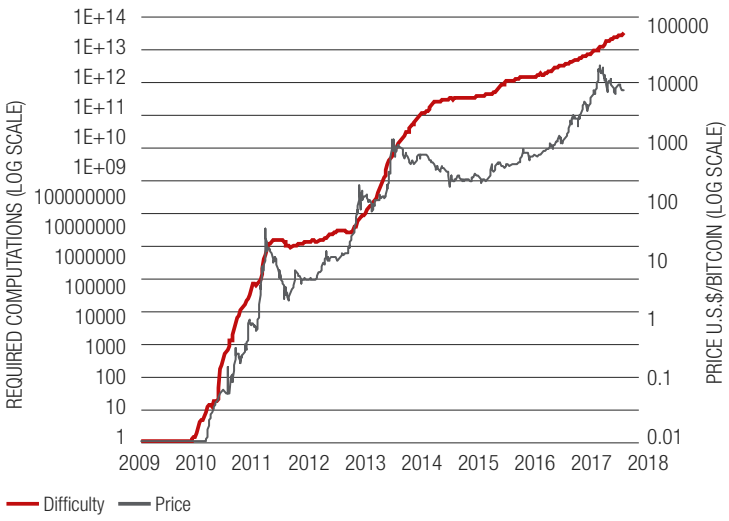
Most importantly, for an asset to function economically as a medium of exchange, it must depreciate slowly over time – something that is impossible with a fixed supply. That loss of value is precisely what makes them useful. Without the fear of inflation, holders of currency tend to hoard rather than spend it. This is why most major central banks, such as the Federal Reserve, European Central Bank, and the Bank of Japan, for example, have set modest inflation targets of 2%, as suggested back in the 1960s by Professor Milton Friedman. The inflation target creates a disincentive to hoard the currency, since hoarding a currency depresses economic growth and creates financial instability. The Japanese yen, the one fiat currency that has experienced deflation over the past few decades, is a case in point. Far from being a virtuous store of value, the Japanese deflation produced a depressed, underperforming economy that the Bank of Japan is desperately trying to turn around with a colossal quantitative easing program four times bigger than that undertaken by the Federal Reserve or European Central Bank, relative to the size of the Japanese economy.

3. A DEEP DIVE INTO BITCOIN SUPPLY THROUGH A STUDY OF THE ECONOMICS OF COMMODITIES

Bitcoin is "mined" by computers solving cryptographic math problems. In exchange for solving the problems, miners receive bitcoins. Those math problems grow in difficulty over time, increasing the required computational power required to solve them. This in turn drives up the equipment, and especially electricity, cost of producing bitcoins. One needs more and more computers and to make them run at peak speeds, they must be kept cool.

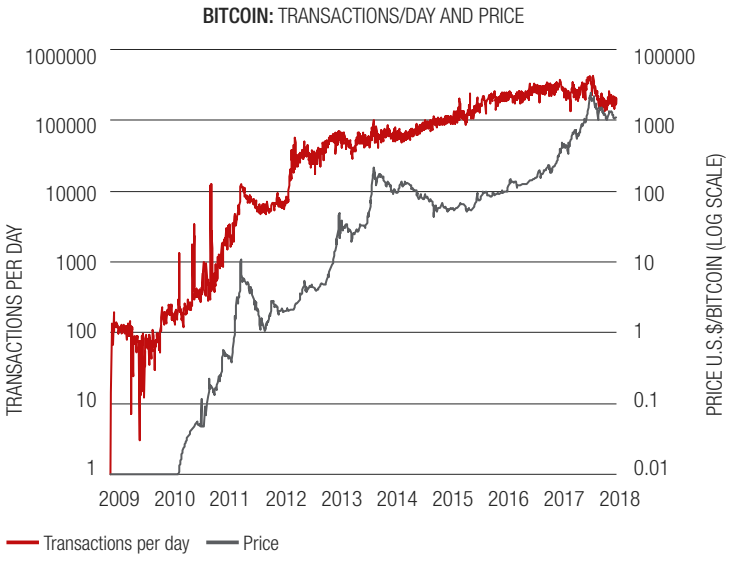
This makes the economic analysis of bitcoin a bit like energy and metals. For example, as of late 2017, the swing producers of crude oil in the U.S. were probably profitable at around U.S.$40 per barrel. Above that price, there are incentives to add to production. Below that price, the incentives are to curtail production. Similar to bitcoin, the difficulty of extracting energy from the earth has increased substantially over time.

Figure 4: Bitcoin mining difficulty and price



Sources: https://bit.ly/2td8ref and https://bit.ly/2OevvaO

Figure 5: Does bitcoin volume drive price?



Sources: https://bit.ly/2td8ref and https://bit.ly/2mudPpK

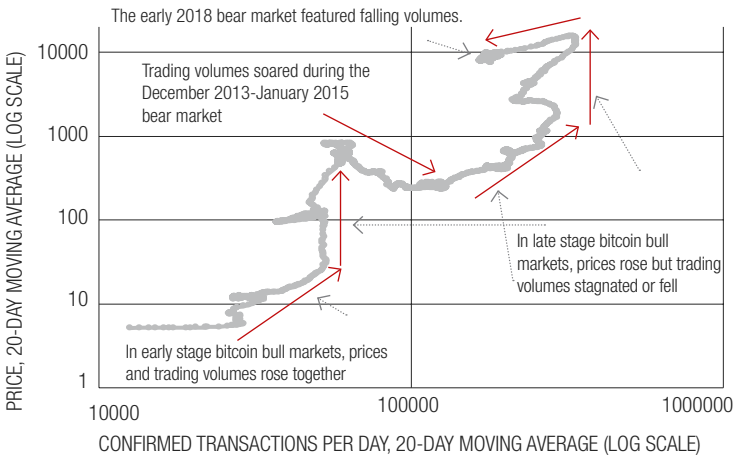
For instance, humanity went through the easiest oil supplies located near the surface many decades ago. Now marginal supply increases come mostly from fracking deep under the ground, from offshore drilling, or from oil in remote, difficult to access locations. In the second half of the 19th century, when oil was first produced in large quantities, one unit of energy invested in oil extraction produced around 150 units of energy. By the 1970s that was down to around 30 units, and that ratio fell to around 15 by 2000. By 2020, this ratio will have probably dipped below 10. This has been a factor in driving oil prices higher. In the energy industry, it is widely assumed that the marginal producers have a cost of production near U.S.$40 per barrel. It bears mentioning that oil prices rarely traded at U.S.$40 until about 2005, when they rose above that level and have only occasionally looked back.

For metals like copper, gold, and silver, there are two numbers to watch: the cash cost and all-in sustaining costs. Cash costs give one a sense of price levels at which producers will maintain current production. All-in sustaining costs give one a sense of what current and anticipated future price levels will be necessary to incentivize additional investment in future production. For example, for gold, cash-cost for mine operators averages around U.S.$700 per ounce while the all-in sustaining costs are around U.S.$1,250 per ounce (Figure 3).

What is interesting for gold, silver, and copper is that after their prices began to fall in 2011, it squeezed the profit margins of operators, who in turn found ways to streamline their businesses and cut their production costs. The same is true of the 2014-16 collapse in energy prices, which may have lowered the marginal cost of production from U.S.$50 to U.S.$40 per barrel of crude. Like mining metals and extracting fossil fuels, mining bitcoin is also a competitive business. Not surprisingly we see a similar feedback loop between the bitcoin price and mining-supply difficulty – in this case "difficulty" is measured in terms of the number of calculations required to solve the crypto-algorithm to unlock a few more bitcoins in the mining process.

Glancing at Figure 4, it is obvious that as the required number of computations (difficulty) has risen, producing bitcoin has become more expensive. It is not a stretch of the imagination to hypothesize that the exponential rise in the difficulty of mining bitcoin has contributed to the exponential rise in price. True, perhaps, but not the

Figure 6: Relationship between prices and transactions



Sources: blockchain.info/charts (price, transactions per day), CME Economic Research Calculations

complete story. There is another side to this feedback loop. Notice what happened to "difficulty" after the first bitcoin bear market (a 93% drop) in 2010-11. Its inexorable rise came to a two-year long halt until prices recovered. It was only when the next price bull market began in 2013 that "difficulty" began increasing again. A similar phenomenon occurred in the aftermath of the 2013-15 bitcoin bear market (an 84% decline). There too "difficulty" stagnated until prices began their next bull market. Curiously "difficulty" did not stagnate in early 2018 despite a more than 50% drop in bitcoin prices from their December 2017 highs, but there are always lags to be studied and examined in future research.

Our conclusion is that bitcoin supply appears to have at least one similarity with that of energy and metals. When prices fall producers must take measures that cause production costs to stagnate or even fall. While "difficulty" never appears to decline, the cost of computing power has fallen over time by as much as 25% per year. As such, if "difficulty" goes sideways for a year, the actual cost of production probably falls as the amount of energy needed to perform the same number of calculations declines. Just as metals and energy producers find ways to reduce cost after bear markets, the bitcoin mining community appears to do the same.

One last comment on supply before we move on to demand: it has long been rumored that the founding community of bitcoin controls something in the order of 3-5 million coins. If this is true, in theory, higher prices could (and probably would) encourage them to part with their coins in exchange for fiat currencies or other assets. When one takes this into account, bitcoin supply might not be perfectly inelastic in the very short term. A similar phenomenon exists in precious metals. When prices rise, we see an increase in the recycling of gold and silver (secondary supply). What is interesting, however, is that recycling appears to respond to price but does not drive prices. For gold and silver, the only supply that appears to drive price is mining supply. Likewise, if an existing holder of bitcoin liquidates some or all of their holdings, this increases its short-term availability but does nothing to influence its total long-run supply, and in that sense, is more like a temporary inventory adjustment.

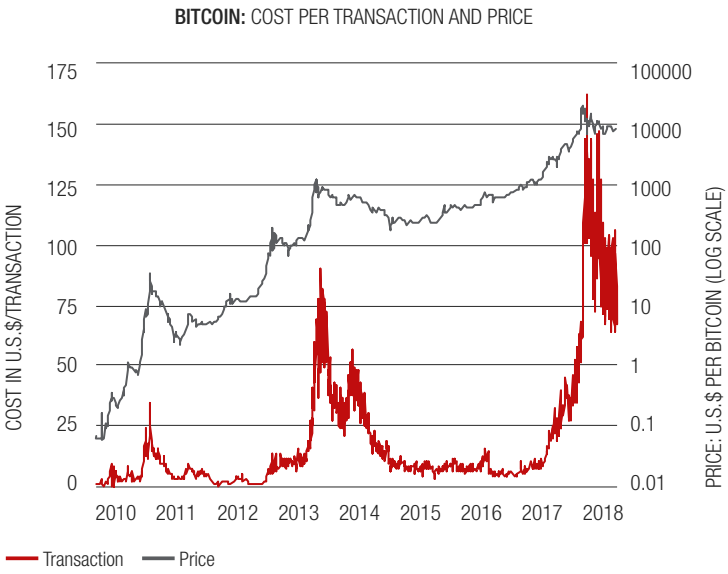
4. DEMAND DRIVERS ARE NOT SO TRANSPARENT

While bitcoin supply is extremely transparent, bitcoin demand is rather opaque. That said, there are a few quantifiable items that we do know about bitcoin demand. First, we have a pretty good idea of the number of bitcoin transactions performed each day. Second, and more importantly, it appears that fluctuations in bitcoin transaction costs play a major role in determining price corrections.

There appears to be a loose relationship between the growth rate of transactions and the rise/decline in price. For example, the number of transactions stopped growing in 2012, about one year before bitcoin's 2013 peak and bear market. It began to rise again in 2014 before bitcoin prices began to recover in earnest but has been stagnating since the end of 2016 (Figures 5 and 6), perhaps foreshadowing the correction in early 2018. What is particularly striking about this correction is that the number of transactions did not rise as prices fell, as they did during the December 2013-January 2015 bear market. During the two previous bull markets, the number of transactions began rising well in advance of the actual rally in bitcoin prices. Towards the end of the two previous bull markets, prices soared as the number of transactions stopped rising.

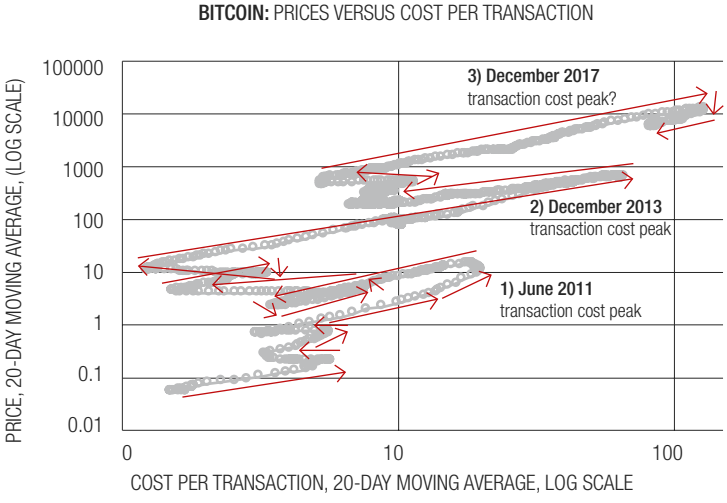
The relationship between bitcoin prices and transaction costs is even more compelling. Trading costs spiked from U.S.$2 to around U.S.$30 per transaction in late 2010

Figure 7: What level of bitcoin transaction costs can the market sustain?



Sources: https://bit.ly/2td8ref and https://bit.ly/2xSRD03

Figure 8: Relationship of prices to transactions costs



Source: blockchain.info/charts (Price, Cost Per Transaction), CME Economic Research Calculations

just before bitcoin prices suffered a 93% collapse. As bitcoin transaction costs subsequently fell, another bull market developed. Transaction costs edged higher in 2012 and then soared to over U.S.$80 by early 2013, which coincided with another collapse in bitcoin prices. By 2015, transaction costs eased towards U.S.$8 and another bull market began. Starting in late 2016, they began to rise again and by early 2018 were up towards U.S.$100-U.S.$150 per transaction (Figures 7 and 8). This third spike in transaction costs may be closely related to the early 2018 correction in bitcoin prices as high transaction costs may have played a role in causing demand for the cryptocurrency to wither at the time.

We are not suggesting that bitcoin prices are a function of trading costs or vice versa; however, there is an association between the two with mutual feedback loops. When bitcoin prices rise, eventually transaction costs appear to rise as well. When transaction costs reach levels that market participants can no longer bear, the price of bitcoin often corrects. A decline in prices puts downward pressure on transaction costs which, at least in the past, allowed for another bitcoin bull market once they had corrected to lower levels.

If stagnating numbers of trades and rising transaction costs do in fact play a role in provoking bitcoin price corrections, then one might hypothesize that a given correction might last until transaction costs fall and the number of transactions begins to rise again. The spike in transaction costs in early 2018 and the sharp (more than 50%) decline in the number of transactions being recorded in bitcoin during the same period led to successive rounds of bitcoin price drops.

5. INCENTIVES, BITCOIN FORKS, AND ALTERNATIVE CRYPTOCURRENCIES

When one thinks of incentives and reward structures, one might want to analyze some parallels with how shareholder value is created. This is tricky with bitcoin. Bitcoin certainly does not fit the definition of a company. It has no board of directors, no balance sheet, no income statement, and no cash flow statement. That said, bitcoin does have a couple of features that need to be understood in the context of incentive structures. And, this adds a little more complexity to the supply analysis as well.

Miners and transaction validators receive rewards in bitcoin. One can see a corporation's shares as an internal currency used to compensate and motivate employees, aligning their interests with those of the organization. To that end, the number of bitcoins in existence is comparable to the "float" of a corporation – the number of shares issued to the public.

When bitcoin forks into a new currency, such as Bitcoin Cash, the move can be analyzed in a manner comparable to a corporate action, such as a spin out. In a spin out, a corporation can give each of its shareholders new shares in a division of the firm that is being released to the public as separate and independent entity. Likewise, when bitcoin most recently forked, the owner of each bitcoin received one unit of Bitcoin Cash, a new and separate cryptocurrency.

> "While bitcoin supply is extremely transparent, bitcoin demand is rather opaque."

In a sense, bitcoin could be viewed as a reference index on the cryptocurrency space more generally. Many new alt-coins, in addition to copying bitcoin's technology, are more easily purchased via bitcoin than they are by using fiat currencies. Bitcoin's central role in this ecosystem makes its price a bit like an index on the health of the entire ecosystem itself. Not surprisingly, the prices of other cryptocurrencies like Ethereum and Ripple are highly correlated with bitcoin when seen from a fiat currency perspective.

A quick diversion back to supply is useful here. The existence of forks in bitcoin serves to modify some of our intuitions on supply. That is, while bitcoin's supply is fixed, the supply of cryptocurrencies is not. Indeed, rising bitcoin prices incent bitcoin forks. This makes a lot of sense, but it does complicate the analysis as it is a reminder that one should not look at bitcoin in isolation but as an anchor for the whole cryptocurrency space.

6. ECONOMIC DESTINY OF BITCOIN

Even if bitcoin fails to replace fiat currencies, it will not necessarily be without long-term economic impact. One possible result of the development of cryptocurrencies is that central banks may one day decide to issue their own distributed ledger currencies, as Venezuela is struggling to attempt to do today with the launch of the "petro." Former Fed Chairman Alan Greenspan once compared making monetary policy to driving a car guided only by a cracked rearview mirror. Even now, important policy decisions must be based upon imperfectly estimated economic numbers that are weeks or months old by the time they become available. In 2018, economic policymaking is still a vestige of the 20th century.

Blockchain technology has the potential to allow policy makers to issue their own cryptocurrencies that will give them real time information on inflation and nominal and real GDP. It will not allow them to peer through the front windshield into the future but at least they can look into the rearview mirror with much greater clarity and see out the side windows of the monetary policy vehicle. This could allow them to create the amount of money and credit necessary to keep the economy growing at a smooth pace more easily than they do today. Switching off the gold standard vastly reduced economic volatility and improved per capita economic growth. Moving to blockchain-enhanced fiat currencies could further reduce economic volatility and, ironically, enable further leveraging of the already highly indebted global economy as people find ways to use capital more efficiently. More broadly, crypto-inspired investments could bring about new technologies that we cannot yet imagine.

Investors who are buying bitcoin are presumably hoping to find someone to sell to at a higher price. That said, there is more to bitcoin economically than just the theory of the greater fool. As more people bid up the price, the difficulty of solving bitcoin's cryptographic algorithms increases. This, in turn, is driving up investment in more powerful and faster computing technology of both a traditional integrated circuit and non-traditional variety. Indeed, solving cryptographic problems may be one of

the first tests facing quantum computers. Finally, we note that investors in bitcoin and its peers are mainly out to make profits and not to finance or subsidize the development of distributed ledgers nor more powerful computers. As such, we caution against linking topics such as the future of blockchain to performance of bitcoin.

The bottom line is:

- Bitcoin supply is highly inelastic; and as with commodities, inelastic supply increases volatility.
- "Difficulty" of mining bitcoin math problems and its price are in a feedback loop, where "difficulty" is a major driver of price, and price also influences "difficulty".
- Transaction volume may influence price trends, and rising transaction costs are a risk indicator for bitcoin.

TRADING BRICKS FOR CLICKS: HONG KONG POISED TO LAUNCH ITS VIRTUAL BANKS

ISABEL FELICIANO-WENDLEKEN | Managing Principal, Head of Digital, Capco Hong Kong
MATTHEW SOOHOO | Consultant, Capco
DOMINIC POON | Consultant, Capco
JASMINE WONG | Consultant, Capco
ANTONIO TINTO | Principal Consultant, Capco

ABSTRACT

Hong Kong's banking regulator, Hong Kong Monetary Authority (HKMA), published a Guideline on authorization of virtual banks, which grants banking status to financial institutions without the need for physical branches. This move is geared to usher Hong Kong into the "new era of smart banking" and capitalize on the significant opportunities brought about by the convergence of banking and technology. The banking sector is set to play a more active role in the semi-autonomous Chinese territory's grand ambitions to transform itself into a "smart city," amidst the backdrop of intensifying regional and global digitization. This article outlines five critical areas for consideration for virtual banks aspiring to succeed in the Hong Kong market: 1) customer focus, 2) continuous product innovation, 3) partnerships and cooperation, 4) the role of the regulators, and 5) new and existing infrastructure. Taking stock of these areas will enable new entrants to stand out amongst the long-established banking giants in the city's hyper-competitive banking industry.

Virtual banking is no longer a buzzword or futuristic concept. Hong Kong's banking regulator, Hong Kong Monetary Authority (HKMA), published a Guideline on authorization of virtual banks,[1] which grants banking status to financial institutions without the need for physical branches (i.e., all transactions will be conducted online). This move is geared to usher Hong Kong into the "new era of smart banking" and capitalize on the significant opportunities brought about by the convergence of banking and technology. It is also part of the semi-autonomous Chinese territory's grand ambition to transform itself into a "smart city," amidst the backdrop of intensifying regional and global digitization.

This guideline broke free from the long-established three-tier banking system in Hong Kong, namely: licensed banks, restricted license banks, and deposit-taking companies. 29 virtual bank license applications were filed as of the 31 August, 2018 deadline.[2] This is a compelling figure considering there are only 22 locally incorporated banks in Hong Kong to date.[3] It is emblematic of the ongoing efforts of large-scale financial institutions to compete head on with the new fintech entrants, yet they are

[1] https://bit.ly/2zEtlUg
[2] Yiu, E., 2018, "Hong Kong draws 29 applications for virtual bank licences," South China Morning Post, August 31, https://bit.ly/2wAwgkm
[3] https://bit.ly/2Nb6FUw

increasingly being outcompeted in an effort to capture the new wave of digital consumers. New all-digital banks are now taking the lead in the fight for banking supremacy. With digital solutions at the core of banks' business strategy, coupled with the willingness to push the boundaries of banking, aspiring virtual banks can take advantage of Hong Kong as an international financial market that is characterized by its proximity to China, stable technical infrastructure, low tax regime, and sound regulatory framework.

This article outlines five critical areas for consideration for virtual banks aspiring to succeed in the Hong Kong market: 1) customer focus, 2) continuous product innovation, 3) partnerships and cooperation, 4) the role of the regulators, 5) new and existing infrastructure. Taking stock of these areas will enable new entrants to stand out amongst the long-established banking giants in the city's hyper-competitive banking industry.

1. SPOTLIGHT ON THE CUSTOMER

Understanding your customer is crucial and fundamental in forming any business strategy. For industries that rely on a continuous flow of new customers as a primary revenue stream, and banking is such an industry, any new entrant must have a customer-centric approach that puts the customers' needs ahead of anything else and relentlessly focuses on how their lives can be improved.

Hong Kong has a long history of banking, housing such traditional giants as HSBC, Standard Chartered, and Bank of China. Going head to head with these banks may be a losing endeavor with an insurmountable hill to climb, but there are ways for aspiring virtual banks to take advantage of the gaps left by these giants. Every bank should aspire to serve all customers and find ways to obtain their business. But focusing on customers whose banking needs have not been fully satisfied by traditional banks may be a more strategic approach. Even with its storied history in banking, only 53% of Hong Kong residents are satisfied with their banking services, as compared to 88% in the U.S. and 78% in the U.K.[4] This suggests that there is an opportunity to adapt banking offerings to fit customers who continue to look for convenience, faster payment capabilities, and lower fees – among other improvements. New entrants have the potential to reach different market segments that have not been fully served by the incumbents.

1.1 The unbanked

This is a key segment that virtual banks in Hong Kong are looking to capture. Customers that have traditionally been unable to find products that suit their needs within large banks make up a sizeable percentage of Hong Kong's market. Catering to the needs of this customer group will boost financial inclusion and provide this segment with lower cost services, in line with HKMA's guidance on not imposing any minimum account balance requirements or low-balance fees. Virtual banks' use of new technology has given them the advantage of being able to reach the far corners of Hong Kong's broad customer spectrum.

As the HKMA sets its sights on a new smart banking era for the region, its guidelines have been designed to promote financial inclusion. Under the auspices of these guidelines and with new technologies at their disposal, virtual banks can develop products to help them reach new customer bases. Virtual banks would be smart to focus on this segment if they wish to compete in Hong Kong's banking market. The most obvious target amongst the unbanked is low-income customers who have not had the means nor the access to traditional banking products. One requirement that the HKMA has imposed to promote inclusion is preventing virtual banks license applicants from imposing minimum balances and low-balance fees.[5] Where traditional banks cannot afford to retain these low balance customers due to personnel, rental, and other overhead costs, virtual banks have some latitude to take on the additional risk due to lower operating expenses. The regulator further adds that, "The HKMA would like to point out that a key objective of introducing virtual banks in Hong Kong is to help promote financial inclusion by leveraging on these banks' IT platforms that would lower the incremental cost of taking in additional customers."[6] As virtual banks can avoid Hong Kong's notoriously high property costs, they have a chance to focus on these customers. With technology at the base of their business model, the cost to acquire new customers would be significantly less.

The upward trend of financial inclusion is not limited to allowing the unbanked to open accounts. There are other categories of customers that can help a new bank take its place in the market. Using technology to provide

[4] Fergus, G., 2018, "Sink or swim. Hong Kong lenders must get smart with virtual banks as fintechs snap at their heels," South China Morning Post, February 19, https://bit.ly/2zEnu77

[5] https://bit.ly/2zEtlUg

[6] Hong Kong Business, 2018, "Virtual banks are no longer a thing of the future in Hong Kong," May 31, https://bit.ly/2xSXa7Z

customers with typical banking services such as transfers and remittances is also a way of improving financial inclusion and breaking into the market. In addition, new and emerging technologies can simplify processes and further lower the transaction costs.

A case in point is the use of blockchain technology: China's Ant Financial Services intends on offering remittance services for almost no cost. Hong Kong has 200,000 Filipinos working as domestic helpers who transfer an average of HK$4.4 billion a year to friends and family in the Philippines, contributing towards the third largest remittance market in the world.[7] In view of the anticipated increase in the number of foreign domestic helpers doubling to around 640,000 in 50 years, virtual banks can create a sizeable customer base by being the market leaders in this particular segment. Hong Kong banks and remittance outlets charge around HK$18 per transaction in addition to commissions and currency conversion fees. With the Alipay blockchain, these transfers can be done at a competitive exchange rate, in real time, and at much lower prices than traditional services. Aspiring virtual banks will continue to use technology to drive their business strategies, create new digital solutions, and new ways of delivery to reach customers that were previously unserved. Technology will fuel digital products and virtual banking with financial inclusion and profitability accelerating in lockstep.

1.2 Generation next: Z

Challenger banks and their micro-lending products can appeal to Generation Z (born from 1995 onwards) customers, many of whom are students with limited credit history. Microbusiness startup loans and self-employment loans could aid this generation launch new companies and deliver their ideas. Self-enhancement loans would provide underprivileged young people with the means to fund tertiary education and acquire additional skills in the increasingly competitive Hong Kong market.[8] This can be groundbreaking for Hong Kong's younger generation who find it increasingly difficult to gain the confidence of larger banks and obtain loans to jumpstart their businesses, education, and life in general. While major banks hold over 95% of the prime plus and super prime loans belonging to older generations, they hold less than half of those issued to Generation Z (Table 1).[9] This can be a potential segment for virtual banks to focus on in Hong Kong.

The low number of unsecured loan accounts is probably due to the banks becoming more stringent with their qualification criteria. Traditional banks are intentionally excluding younger borrowers, opting to withhold loans until their careers are more established. Fintechs and challenger banks can now step in to service this segment. Only one in six Hong Kongers aged between 18 and 22 are currently in the banking ecosystem.[10] As the unbanked/underbanked youngsters look to obtain their first loans or look for their first payment channel, they will soon have the option to turn to fintechs and virtual banks rather than the traditional lenders.

1.3 The digitally savvy millennials

What do millennials want? As Hong Kong's smart banking era arrives, virtual banks applying for licenses want to know how they can attract the city's new, digitally savvy customers and ensure they can become a big part of making it a smart city with respect to banking. For

Table 1: Hong Kong unsecured loan accounts distribution

	SILENT (Born 1945 and earlier)	BABY BOOMER (1946 to 1964)	GEN X (1965 to 1979)	MILLENNIAL (1980 to 1994)	GEN Z (Born 1995 onwards)
SUBPRIME	52%	25%	28%	27%	7%
NEAR PRIME	78%	58%	65%	63%	26%
PRIME	95%	86%	88%	82%	31%
PRIME PLUS	99%	96%	95%	86%	31%
SUPER PRIME	100%	97%	96%	86%	50%

Source: TransUnion HK

[7] Soo, Z., 2018, "Ant Financial taps blockchain technology to offer cheaper international money transfers," South China Morning Post, June 25, https://bit.ly/2NbrKhM
[8] https://bit.ly/2R9TFlu
[9] Transunion, 2017, https://bit.ly/2P2d1Yo
[10] https://bit.ly/2P2d1Yo

the last 10 years, the assumption was that traditional financial institutions were the only avenue through which they can satisfy their banking needs. But now the largest portion of the city's workforce of nearly 1.7 million,[11] millennials, and their taste for digital and quick solutions, have started to change the way banks construct their offerings.

From broadband internet to instant food delivery, the need for seamless, efficient, and fast services has never been greater. To win over consumers, any prospective virtual bank needs to continuously focus on and prioritize providing seamless transactions.

> "With millennials increasing in prominence and wealth, the entire banking industry is striving to rapidly adjust their offerings and business strategy to meet their demands."

Payments is perhaps the most basic and prevalent financial interaction with the masses, and yet for the longest time it saw minimal innovation.[12] This was a major pain-point for millennials who are used to sending everything from photos to documents electronically – having to withdraw physical cash or obtain account details to securely transfer money for something as simple as lunch.[13] In other developed markets, these pain points led to creation of e-payment gateways, such as PayPal, and digital wallets, such as Venmo. Traditional banks in the region were quick to follow suit, launching P2P payment platforms, such as Zelle and PayMe. Ultimately, it is efficiency, not brand names, that determines the platforms that millennials will trust to make their transactions on. Virtual banks applying for licenses in Hong Kong would be wise to invest in these types of solutions.

As millennials increase in number and influence throughout Hong Kong, their needs are increasingly important. Unlike previous generations, millennials are hooked on the aesthetic: catchy taglines and jingles are not as effective on this segment. Visual details are paramount for the new generation, as design has become their most desired stimulus. Preference is given to visual content with digital channels, such as social media feeds, apps, chats, and blogs being of utmost importance. Developing live chat interfaces, AI-based chatbots, and even online tutorials with valued content will pay dividends for challenger banks looking to sell their services to younger consumers in Hong Kong. Anything less will only result in frustration, causing them to seek alternative options.

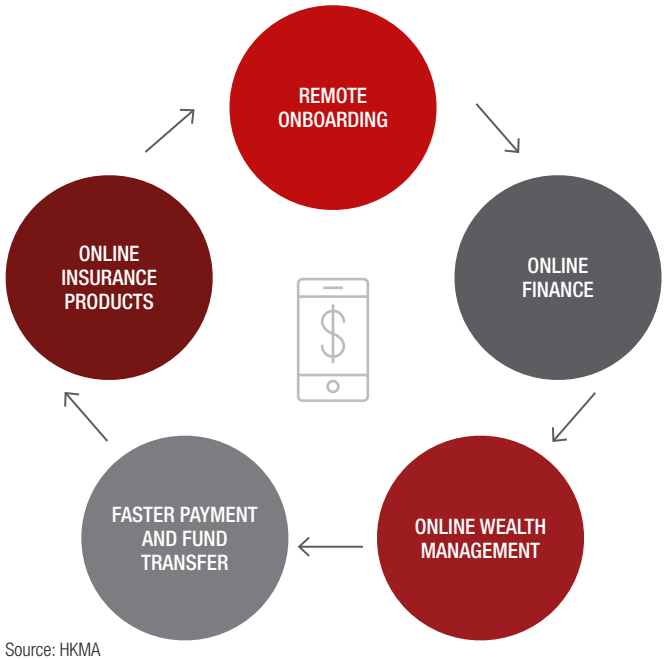
Unlike Baby Boomers and Generation Xers, millennials are used to sharing private information through a multitude of different social media platforms and digital streams. Their attitude differs greatly from their predecessors in that easy registration and services with open ID are preferred, as compared to traditional secure forms. Transparency has become a 'must have' for younger consumers. As a result, they are willing to sacrifice privacy to receive maximum level of service and efficiency. According to a salesforce.com study, 61% of people do not mind sharing their personal information with businesses, hoping that it will make a product or service interaction more personalized. Also, 58% of millennials indicated a willingness to share this information to get more accurate recommendations related to their interests, as compared to 41% of Baby Boomers.[14] The younger generation is already sharing most of their life voluntarily, so they see little harm in offering up some more if it helps simplify their user experience. Although new challenger banks should not be afraid to offer this privacy for data swap, this is a two-way street. With increased access to their personal information, millennials expect complete honesty, transparency, and personalization. Privacy and security should be further safeguarded given the free flow of information. But as the idea of privacy becomes increasingly foreign in this era of smart banking, finding the right balance between security and data collection is something that any virtual bank will need to focus on.

With millennials increasing in prominence and wealth, the entire banking industry is striving to rapidly adjust their offerings and business strategy to meet their demands. Various banks are taking different approaches, but all of them are mobilizing in preparation to receive what they know will be the main driver of the economy in the near future.

[11] Census and Statistics Department, The Government of the Hong Kong Special Administrative Region, Population Estimates, https://bit.ly/2IrEDmU
[12] Tan, S., 2017, "The future of Millennial banking," Marketing Interactive, July 18, https://bit.ly/2R9pb2T
[13] Tan, S., 2017, "The future of Millennial banking," Marketing Interactive, July 18, https://bit.ly/2R9pb2T
[14] Kreger, A., "How to design a millennial bank," Financial Brand, https://bit.ly/2mI5X78

For example, Standard Chartered announced that it plans to apply for a virtual bank license with the HKMA.[15] As part of its goal to enhance client experience and appeal to Hong Kong's digitally-savvy and growing millennials population, it plans to launch a digital-only bank that is run completely independent of its traditional banking operations. Although they are currently retaining their physical branches and presence, staff will be required to learn digital skills and teach customers to use digital solutions for simpler transactions.[16] Other traditional banks, such as HSBC, DBS, and Citibank, are taking a different approach. With well-established branch networks and digital banking services, they have decided to expand on their current digital offerings rather than apply for a virtual banking license. Though their methods differ, the goal remains the same. Attract the attention of and gain business from the digitally savvy millennials.

Figure 1: Virtual bank business model



Source: HKMA

[15] https://bit.ly/2NNylEd
[16] Yiu, E., 2018, "Standard Chartered Bank sees virtual bank licence as ticket to new business worldwide," South China Morning Post, August 6, https://bit.ly/2OJD0Ed
[17] Yiu, E., 2018, "Hong Kong prepares to usher in virtual banks, as 60 firms apply to be pioneers in financial revolution," South China Morning Post, July 28, https://bit.ly/2LOMwXK
[18] Yiu, E., 2018, "Hong Kong prepares to usher in virtual banks, as 60 firms apply to be pioneers in financial revolution," South China Morning Post, July 28, https://bit.ly/2LOMwXK

In a world where the most impactful customer segment orders meals, books tickets to distant countries, and calls taxis with a couple clicks on their phone, banks will not only have to compete with each other, but with other services in the consumer's life. To win over millennial consumers, transferring activity into the digital world is not enough. Virtual banks will have to put millennials' needs at the center and double up their efforts to ensure that their solutions provide the highest levels design, expediency, and transparency. The companies with the best understanding of how millennials think and how they operate will be able to find sustained success and obtain customer loyalty.

1.4 Small and medium-sized enterprises (SMEs)

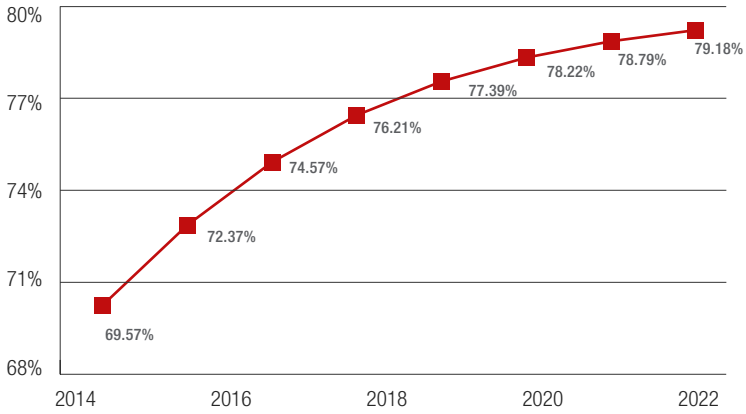
Traditional banks pay an average of HK$1 million per month for a 2,000 square foot storefront and an additional HK$1 million per month for monthly staff costs.[17] Adding in the costs for operational services in the back office, such as accounting, compliance, security, cybersecurity, and technology, one can see why they would have no interest in loans under HK$200,000. This can have the effect of marginalizing smaller businesses, where a loan can take up to six months to be approved.[18] SMEs have struggled to get their businesses off the ground in Hong Kong because of how onerous it is to open a business bank account with traditional banks. These SMEs are the lifeline of any competitive market and are vital to its sustained growth. With the decreasing financial burden of a physical presence, virtual banks can minimize the current operating costs and afford to create new products for customers that were originally seen as unprofitable. This will aid in the development of such solutions such as micro-lending and trade financing, which will allow virtual banks to provide loans to these SMEs and can have the multiplier effect of boosting the economy.

Similarly, overseas companies who are looking to set up entities in Hong Kong encounter a long and arduous process ranging from four to six weeks to open corporate bank accounts, given the KYC requirements. Virtual banks that focus on the SME segment can potentially solve this pain point and offer frictionless onboarding with the use of innovative Identification and Verification (ID&V) and Optical Character Recognition (OCR) technologies. Focusing on responding to the unmet needs of these previously unbanked SMEs, challenger banks can gain a significant advantage over its competitors.

2. INNOVATE TO ACCELERATE

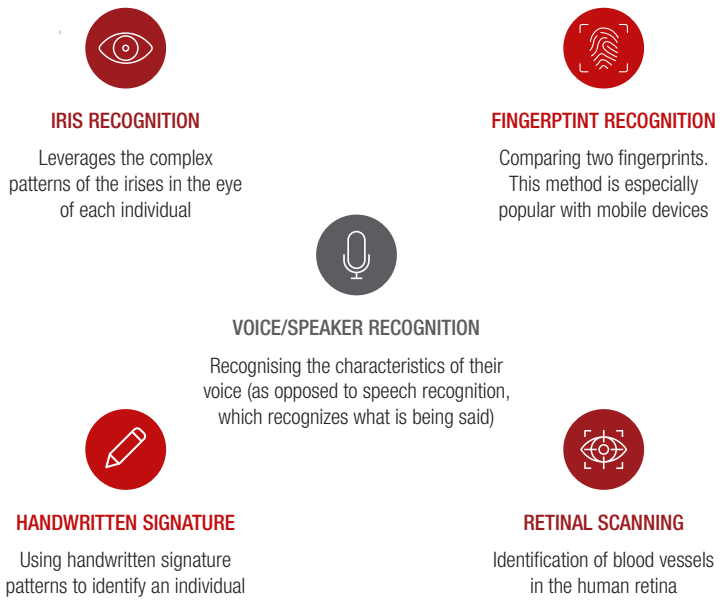
As Hong Kong continues its transition to a 'smart city,' market opportunity for challenger banks increases. These banks must continue to scan the landscape for new ideas to enhance every part of the customer journey and use their technological advantage over large banks to compete for their business.

Figure 2: Hong Kong mobile penetration rate



Source: Statista.com

Figure 3: Biometric authentication methods used by virtual banks



Offering new products and innovations along the journey that target customers embark on and simplifying interaction points will allow virtual banks to position themselves for success well into the future. In this new era of smart banking, devising a way to center their business strategy around enhancing the customer experience across different products and features specific to target segments (see Figure 1) must be paramount if newcomers to Hong Kong's virtual bank market intend to oust the entrenched incumbents.

Hong Kong has the world's 7th highest mobile phone penetration rate, at 76 percent, and yet only 14 percent of them are using those phones to connect to their bank daily; it has become one of the world's largest, untapped markets for tech-focused virtual banks with mobile solutions at the core of their brand (Figure 2).[19] Virtual banks and traditional banks are wrestling for that market opportunity, introducing new products to ensure that their company plays a major part in providing the remaining 86 percent with the means to seamlessly complete daily banking activities from their phones.

Challenger banks must take advantage of recent advancements in biometrics, AI, and blockchain to make improvements to remote onboarding, online financial management, and payment/fund transfers. The introduction of new products for each of these links on the customer journey can help companies capitalize on Hong Kong's growing number of tech-savvy customers and book themselves a spot on Hong Kong's diverse banking industry top table.

Biometrics has become an important part of the digital solutions offered by banks across the world. Advancements in this field have become the norm to revamp the customer onboarding process and provide a higher level of security (Figure 3). Consumers agree that the use of their distinct and measurable characteristics to authenticate use and verify activities is the best way to ensure the safety of their accounts during remote onboarding and throughout the life of their account. Traditional methods of authenticating have been consistent sources of security hazard and user interface/user experience (UI/UX) pain.[20] It is imperative for virtual banks to utilize the selection of different biometric authentication methods that are now available and being leveraged by traditional banks.

The methods presented above[21] are currently being used by virtual banks in other advanced economies,

[19] Gordon, F., 2018, "Sink or swim. Hong Kong lenders must get smart with virtual banks as fintechs snap at their heels," South China Morning Post, February 9, https://bit.ly/2zEnu77
[20] Live Bank, "A new definition of security. Biometrics in digital banking," https://bit.ly/2xXkzEr
[21] Goode, A., "Biometric trends for 2017" 2016, Veridium, December 15, https://bit.ly/2QhWJur

bypassing the need for customers to visit a branch to sign paperwork and create accounts physically. Aspiring virtual banks must take advantage of biometrics as a means of remote onboarding and security to appeal to customers and capture a profitable market share. By simplifying these processes using technology, they contribute to the seamless customer experience. However, careful consideration of Hong Kong's regulatory frameworks and guidance on customer due diligence remain a prime imperative.

Enhancing the customer journey does not stop with onboarding. Providing customers with options to help them manage and monitor their finances and wealth with digital, mobile solutions has become a trend that challenger banks must take advantage of. Using AI to offer personal finance management (PFM) and savings assistants within their app can help any new bank lure in new customers and improve the overall customer engagement. Giving customers insights into their spending habits, views of their overall financial portfolio, and allowing them to act on this information can be a pull for future users.

Recent surveys indicate that 43% of Hong Kong millennial customers born post-1990 save less than 10% of their monthly income, with one in four spending the entirety of their salary each month.[22] Furthermore, a third of the post 1980s and 1990s millennials do not even remember how they spend their money.[23] In one of the most expensive cities in the world, virtual banks that offer AI-based tools to help people improve their spending habits and organize personal finances could be a great source of differentiation. By committing to meeting the financial needs of an increasingly mobile and digital workforce, aspiring banks can help attract and retain customers.[24] Hong Kong customers lack the tools needed to manage their finances properly. If new banks can use technologies such as AI and machine learning to introduce efficient, user-friendly tools, they can take advantage of the growing market opportunity. New products and innovations driven by the desire to improve the customer journey will help new challenger banks succeed in the hyper-competitive Hong Kong market.

In a world that has become increasingly connected through digital and online services, a sure way for a bank to make its mark and boost its services and capabilities is in the field of remittance and fund transfers. Digital wallets are not new, but the functions included in each company's wallet and use of blockchain to power remittance services are of increasing importance. The ability for customers to send money and make payments under various circumstances is of great

Table 2: Hong Kong's payments and fund transfer landscape

FUNCTION	WECHAT PAY	ALIPAY	PAYME	TAP & GO	TNG	O! EPAY	APPLE PAY, GOOGLE PAY, UNIONPAY QUICKPASS
CREDIT CARD BINDING	✓	✓	✓	—	—	✓	✓
TOP-UP	—	—	✓	✓	✓	✓	—
ONLINE PAYMENT	✓	✓	—	✓	✓	✓	✓
MERCHANT PAYMENT	✓	✓	✓	—	✓	✓	✓
P2P TRANSFER PAYMENT	✓	✓	✓	✓	✓	—	—
IN-PERSON MONEY TRANSFER	✓	—	—	—	—	—	✓
GLOBAL MONEY TRANSFER	✓	—	—	—	✓	—	—
BILL PAYMENT	—	✓	—	—	✓	—	—

Source: moneyhero.com

[22] Fintech News Hong Kong, 2018, "Hong Kong startup puts all your banks in a single app," March 8, https://bit.ly/2FpUgZM
[23] Fintech News Hong Kong, 2018, "Hong Kong startup puts all your banks in a single app," March 8, https://bit.ly/2FpUgZM
[24] Finextra, 2018, "Hong Kong digital bank Neat scores $2m funding boost," July 5, https://bit.ly/2IVH0Er

value. As cash becomes less frequently used in Hong Kong, improving the ways that users can remit, transfer funds, and pay has become a critical component of the customer expectations.

Digital wallets with the capacity to complete instant money transfers, cross-border peer to peer transfers, bill payment, global transfers, and top-up functions are 'must haves' for any of the Hong Kong millennial customers that challenger banks seek to attract (Table 2). One local, digital wallet startup, TNG Wallet, has been able to break away from the pack in recent years. With over 600,000 customers, comprising of 9% of the city's population,[25] TNG Wallet is an example of how a newcomer virtual banks can grab a sizeable market share in a landscape filled with giants. By offering a multitude of different options in its wallet and continuously adding to its repertoire of digital services and solutions, it has provided a blueprint for aspiring virtual banks to follow. AliPayHK already utilizes blockchain technology and introduces a variety of fund transfer options that will be crucial for increasing their customer base and profits. Other aspiring banks can follow suit by focusing on, and considering technology-based products that enhance their customers' experiences and allow them to make payments and transfers in a variety of circumstances.

Hong Kong's burgeoning fintech and digital banking industries show promise. However, there is no one blueprint to ensure success, and there is no single product or innovation that can bolster an entire virtual bank. Newcomer challenger banks must be diligent in their research to fully understand the products that customers value. Enhancing the whole customer experience and introducing products that improve parts of the customer's virtual banking journey can lead to sustainable success and garnering customer loyalty. From customer onboarding, peer to peer transfers, to instant money transfers, Hong Kong's customers demand more from their banks. Tremendous strides have been made in biometrics, AI, and remittance technology. Focusing on these advancements in emerging technology as a guide to improving the customer journey must be considered for all aspiring virtual banks.

3. PARTNER TO WIN

Building a banking business from the ground up in this highly saturated city requires extensive analysis of the various stakeholders, which acts as the foundation for any challenger firm looking to identify a viable proposition, set up effective strategies, and leverage any existing market capabilities. Furthermore, an understanding that the HKMA plays a decisive role through its smart banking frameworks and initiatives will only support a challenger firm to thrive competitively in the city.

3.1 The rise of partnerships and "coopetition"

Partnerships between banks and fintechs are springboards for innovation, speed, scale, and access to large customer bases. It is becoming increasingly prevalent in the recent years, with notable collaborations between the likes of Standard Chartered and AliPay,[26] as well as Bank of China HK and Tencent.[27] Traditional competitors are teaming up to form alliances with various objectives in mind, such as jointly setting up fintech labs and accessing more customer bases together.

As such, this creates the perfect storm for new financial propositions to be offered to a wider audience. For large banks, partnerships are logical, as they can tap into challenger firms who are able to offer them a new technology or capture a new segment of customers with whom they historically could not engage. For instance, the relatively young, tech-savvy, and affluent demographic, have demonstrated that they are more open to using non-traditional banking institutions.[28] Likewise, they also possess an overt interest in new digital offerings, such as mobile apps, mobile P2P payments, and smart wearables.[29]

The increasing shift towards strategic partnerships gives financial institutions the opportunity to capture a customer base who have historically had relatively little interactions with traditional institutions. Overall, collaboration provides mutual benefits. From the perspective of a challenger bank hoping to break into the Hong Kong market, a partnership with a larger bank enables the former to tap into their knowledge and expertise regarding local demographics and regulatory knowledge and risk control.

[25] Fintech News Hong Kong, 2018, "8 future fintech unicorn startups in Hong Kong to watch in 2018," January 5, https://bit.ly/2Ay0ow7
[26] https://bit.ly/2Ooa6Mi
[27] Zhang, M., 2017, "Big banks strike partnerships with technology companies as part of fintech wave," South China Morning Post, August 27, https://bit.ly/2OkizQG
[28] Anand, V., "Today's Banks Need a Millennial Banking Technology Framework," 2018, Finextra, January 19, https://bit.ly/2n3LngP
[29] Nielsen, "Smarter digital city," in collaboration with Google, https://bit.ly/2OWe9wl

3.2 What are others doing right?

New fintech and banking solutions are thriving in the market. Extensive analysis of the edge and deficiency of the competitors relative to the firm's strategy is essential in gaining a market advantage. A review of the profiles of 29 license applicants and the more extensive global offering confirms that plenty of potential could be unleashed as the majority are specialists in payments and online lending, rather than end-to-end digital-only virtual banking.

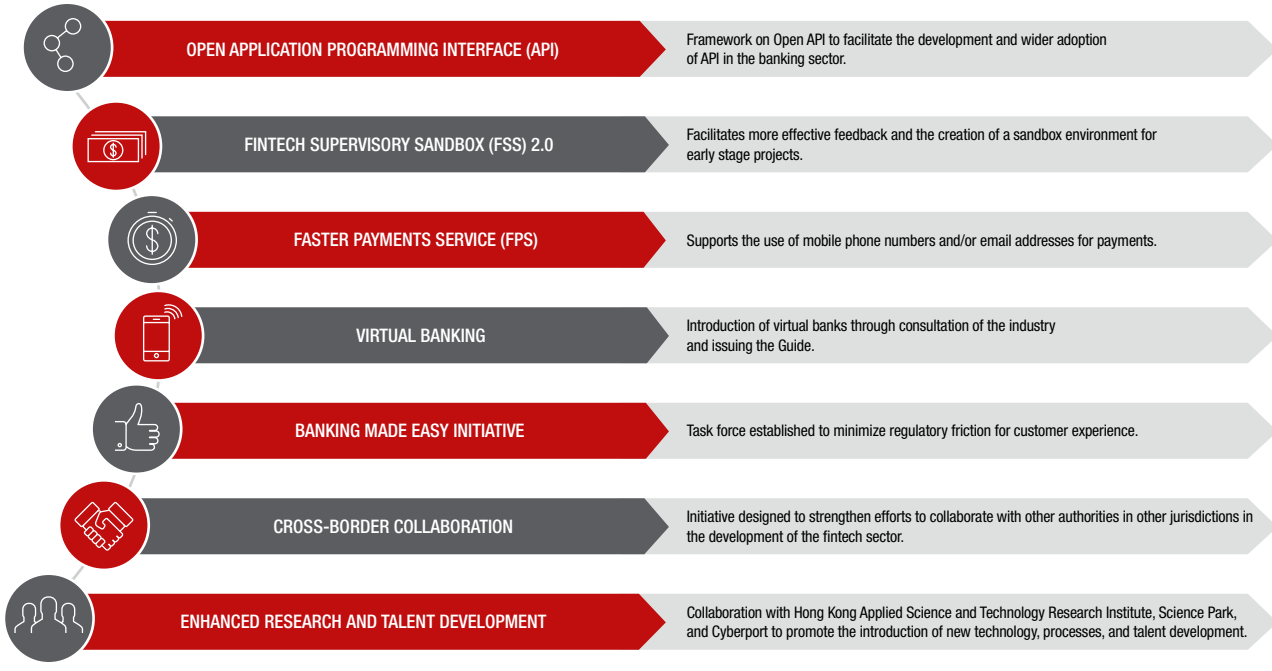
There is much to learn from existing Hong Kong mobile apps, such as HSBC PayMe, which offers the ability to pay friends and peers through the app seamlessly. Apps in advanced markets such as HK's Neat, the U.K.'s Revolut, and Germany's N26, which provide the ability to aggregate financial views by customizable measures, are also interesting to look at (Table A1, in the Appendix). Looking at best practices from first mover players and acquiring the best-fit solutions that cater to Hong Kong consumers will enable virtual banks to accelerate their growth in the market and develop targeted solutions amidst the crowd of predominant players.

4. WORKING WITH THE REGULATORS

The government and HKMA's strong support to establish the digital ecosystem for the banking industry in Hong Kong is reflected in new initiatives and investments. Hong Kong's Financial Secretary announced that the 2018-2019 budget will dedicate over HK$50 billion towards developing the city's IT infrastructure.[30] As such, official bodies play a decisive role in supporting innovation to maintain the competitiveness of financial hubs.

The successful implementation of innovation, such as the Stored Value Facilities (SVF) wallets in Hong Kong was largely due to the HKMA's involvement in the regulation and supervision of the SVF ordinance. The regulatory body was ultimately responsible for granting licenses and support. Similar guidance has been provided to companies looking to establish virtual banks in Hong Kong. The guidance under a dedicated team who have

Figure 4: HKMA's smart banking initiatives – "New era of smart banking"



Source: HKMA, https://bit.ly/2OV9dld

[30] Onag, G., 2018, "HK Budget 2018: Innovation gets a big boost with $50B investment," ComputerWorldHK, March 1, https://bit.ly/2y1CWb6

published the Guideline on authorization of virtual banks[31] provides clear direction and conditions required. With similar levels of guidance to the SVF initiatives, observers are optimistic that the adoption of virtual banks will be as seamless as that of the virtual wallet initiative.

Furthermore, as the virtual banking initiative progresses, HKMA has been proactive in giving further clarity and support to license respondents.[32] An awareness and understanding of guidelines from official government bodies can only serve to support new challenger firms in establishing a Hong Kong presence. As such, it is beneficial for virtual banks to continue to work closely with the HKMA, who have promoted several initiatives to support the era of 'Smart Banking' (Figure 4).[33] For instance, the Faster Payment System (FPS) allows banking participants to integrate the use of email addresses and mobile numbers for payments. Whilst, the Enhanced Fintech Supervisory Sandbox (FSS) 2.0 will allow early pilot trials of fintech products, with the support of regulators.[34]

5. INFRASTRUCTURE BUILD-OUT AND CO-EXISTENCE

One essential step to fully-digitalized banking services is to modernize the IT infrastructure – from the complex legacy systems. The innovative infrastructure uplift can bring value to traditional banks by speedy transformation, cost reduction, agility on compliance, and security improvement. Shifting from a 'wait-and-see' stance, the realization of digital banking offerings has drawn a lot of attention from curious customers. It has resulted in a motivated drive for innovation by governments and incumbent banks worldwide in the last decade. The continued dependency on tightly coupled legacy infrastructure combined with technical lag in Hong Kong has inevitably opened up a gap for challenger banks to proactively address the needs of customers.

Virtual banks enter the ecosystem to fulfill the everchanging needs of the modern consumers and could deliver their needs through Agile ways of working. Although virtual banks do not suffer from legacy issues, they must still ensure that the banking innovation they offer will contribute towards the wider aims of Hong Kong's 'Smart City Blueprint'. Furthermore, to stand out in this city, they must also consider taking on an educational role for the HK public on innovation.

Industry leaders agree that despite the growth in mobile innovations, banks are still far behind and outdated IT systems continue to be one of the greatest barriers for businesses to fully realize the benefits of digital transformation. Although this is largely due to business infrastructure, reliance on outdated IT systems is as much of a workforce mentality as it is the physical IT architecture itself.

While workers see the value of an innovation shift from outdated systems and are often pained by the inefficient and messy processes, lack of knowledge about innovative ways of working prevents them from moving away from the familiarity of legacy frameworks and processes. A recent HSBC survey supports this proposition: reporting that despite recognizing the value for digital innovation, only 28% of the 300 companies polled had plans for adoption.[35] This is due to a lack of understanding and awareness of the digital economy, with 40% of Hong Kong workers believing that they do not have the adequate training for digital and transformational initiatives.[36]

To create a digitally- and customer-minded workforce, there is an overt need for change in terms of workforce behavior and redundant policies relating to legacy systems. Industry experts believe this can be achieved by further investments in digital capabilities, resulting in improvements in customer service, employee experience, and long-term financial benefits for banks.[37] As such, the need to transform businesses and Hong Kong workforce has resulted in the creation of the HKMA's Smart Banking Initiatives and the 'Smart City Blueprint'.

The recognition that the city's IT infrastructure requires modernizing is the first successful step in transforming the relationship between the customer and the bank. The city's goal to improve the digital user experience aligns perfectly with what aspiring virtual banks aim to offer in Hong Kong. Initiatives such as 'Banking Made Easy' aim to reduce regulatory friction and to address inefficient areas in the customer journey, such as onboarding.

[31] https://bit.ly/2NMXsa8
[32] https://bit.ly/2OmqgWD
[33] https://bit.ly/2DChMqd
[34] https://bit.ly/2NaNRET
[35] Chen, L., 2018, "Businesses in Hong Kong slow to adopt digital tech despite growing demand, finds HSBC survey," South China Morning Post, June 21, https://bit.ly/2Oojek2
[36] Chen, L., 2018, "Hong Kong workers feel they are not equipped with the right skills for digital economy: survey," South China Morning Post, June 13, https://bit.ly/2QiVU4B
[37] Asian Banking and Finance, 2017, "Systemic drawbacks drag Hong Kong banks in digitalisation race," October 2, https://bit.ly/2NSDVFI

From the virtual banking perspective, this is the type of customer issue it has always aimed to solve. As mentioned in previous sections, onboarding trends, such as biometrics, are already being harnessed by digital banks such as Revolut and Monzo in the U.K., who have invested heavily in creating a seamless user experience. Virtual banks can also align themselves to other initiatives, such as the Faster Payment Service (FPS), allowing customers to pay using their email addresses or mobile numbers. With the focus on the customer journey and user experience, it is inevitable that challenger banks will become a part of both the smart banking and smart city plans, to make an impact in Hong Kong's hyper-competitive financial landscape.

The innovation journey of incumbent banks will depend on whether they can proactively break away from legacy systems. This will also come down to the bank's ability to educate its workforce, and to further invest in innovation and partnerships. The advent of virtual banks will only serve to further pressure incumbent banks to innovate or seek strategic partnerships with fintech firms. On the flip side, challengers looking to establish a virtual banking presence in Hong Kong should be aware of the prevalence of both the legacy infrastructure and the worker ties to these outdated systems and processes. Virtual banks benefit greatly from lack of constraints of outdated systems and processes. This provides them with an advantage of delivering new propositions to the customer in an agile fashion. However, to spearhead banking innovation in Hong Kong, challenger firms must embrace the government's wider aim of promoting and building the 'Smart City Blueprint'.

6. CONCLUSION

This article highlights five factors any new challenger bank looking to establish a presence in Hong Kong must consider. An understanding of the market's customer segments; an awareness of competitors and potential partners; an acknowledgement of the government's as well as the regulator's role; and the current infrastructural state of the city.

With new digital solutions emerging every day and a multitude of new customers up for grabs, companies are scrambling to get a piece of this customer segment. Customer bases in remote corners of Hong Kong are now within reach due to the advent of blockchain and big data, as well as offering services such as microloans and payments. Newcomers to the banking scene must be diligent and laser-focused when determining which customer segments they will target. The money saved by reducing their physical presence in the world's most expensive real estate market has accelerated the process of financial inclusion, thus removing roadblocks to a broader array of customer demographics. The competition between traditional banks' digital platforms and fintech companies has only grown fiercer as they battle to enhance the customer experience and improve the convenience of banking activities. Supplemented by the initiatives and advocation of virtual banking by the city's regulating bodies, the bar continues to rise across Hong Kong's banking industry.[38] As technology continues to advance and competition intensifies, customers from different segments can potentially reap the benefits. Ultimately, it boils down to the new, aspiring banks taking advantage of these openings in the Hong Kong market.

Virtual banks now have a real opportunity to not only generate financial revenue, but to also more significantly create trust and connect with new customer segments who have historically had little interaction with traditional financial institutions. Likewise, virtual banks must take a proactive stance towards the continued existence of legacy systems and mindset by engaging in aspiring initiatives through knowledge sharing in a growing innovation community. To fully realize this role, virtual banks must continue to work closely with the city's regulators, who continue to play a supportive role, as part of Hong Kong's journey to evolve into a smart-banking smart city. To stand out in this hyper-competitive banking industry, virtual banks must continue to harness new banking innovation to capture the ever-changing needs of the Hong Kong customer.

[38] Olsen, K., 2018, "Virtual banking is set to shake up Hong Kong," CNBC, July 18, https://cnb.cx/2QgQEOR

Table 3: Global competitive landscape – major players

	REVOLUT	MONZO	N26	STARLING BANK	DIGIBANK	BUNQ	ALLY
LOCATION	U.K.	U.K.	GERMANY	U.K.	SINGAPORE	NETHERLANDS	U.S.
KEY FEATURES	• Custom spending limits • Contactless payments • Multi-currency • Dashboard analytics categorizing payments • Intelligent push notifications • Split bills • Worldwide travel insurance	• Spending limits • Contactless payments • Multi-currency • Apple & android pay dashboard analytics categorizing payments • Intelligent push notification • Split bills • Tag company expenses	• Joint account • Contactless payments with Google Pay • Overdrafts • Spending limits • Transaction tagging • Dashboard analytics categorizing payments • Intelligent push notification	• Joint account • Wide range of digital wallets from Apple to Garmin pay • Overdrafts • Spending limits • Weekly to monthly spending insights • Intelligent push notifications • Marketplace (API plugins of financial products)	• QR code scan payment method • Overspending notifications with suggestions and solutions offerings • Budget optimizer to understand behavior and provide recommendations based on synthesized data	• Joint and group bank accounts • Multiple PIN numbers for one card • Children's bank accounts supervised by parents aimed at educating children with a safe and simple way manage money	• AllyCard Control app allows you to decide when and where your card is used • Use of Amazon's personal voice assistant, Alexa, to complete advanced banking activities and transactions
SECURITY FEATURES	• Freeze card • Report lost or stolen • View or change card pin	• Freeze card • Report lost or stolen • View or change card pin	• Freeze card • Report lost or stolen • View or change card pin	• Freeze card • Lock online payments, gambling payments and mobile wallet • View or change PIN • Location based fraud detection	• Embedded soft token for security instead of SMS for one-time passwords	• Set spending limits and block your cards at any time • Use of rotating CVC codes for secure online shopping	• One touch freeze or unfreeze card options
CUSTOMER SUPPORT	• 24/7 chatbot and live support	• 24/7 live support	• 24/7 live support	• 24/7 live support	• AI-based virtual assistant	• Online use forum • Online support	• Ally Assist, a virtual customer service assistant
SIGN-UP/ ONBOARDING JOURNEY	• ID&V check • Facial scan	• ID&V check • Facial scan	• ID&V check	• ID&V check • Video identification	• ID&V check	• ID&V check • Facial scan	• Online application
LOG-IN JOURNEY	• TouchID or passcode	• TouchID or passcode	• Username and password	• TouchID or passcode	• TouchID or passcode	• 4 finger touchless ID authentication	• TouchID or passcode

FINANCIAL AND DATA INTELLIGENCE

CHARLES S. TAPIERO | Topfer Chair Distinguished Professor, Department of Finance and Risk Engineering, New York University, Tandon School of Engineering

ABSTRACT

Data and computers on steroids have partnered to transform finance and reengineer its future. Past conventions have defined the role of data to be a complement to financial theories, providing a testing ground and an estimator of future prices, whether of assets, stocks, or derivatives. Theories of finance (such as the Arrow-Debreu framework, Martingale pricing, risk neutral pricing, etc.), while mathematically and theoretically stimulating, also embed a variety of risks and real financial misconceptions. For example, risk is defined by predictable (future states) events while in real finance, uncertainty primes [Knight (1924)]; prices exist only in the present and so on. Further, while conventional finance is an ex-ante approach to the future, data science is an inverse approach that seeks ex-post to estimate causes or models that explain the data so collected and improve their state of knowledge and know-how by learning through a feedback process. This approach is often structured by terms such as "deep learning," "machine learning," and "artificial intelligence." Thus, one approach is defined by hypothetical theories, while the other is an analytic data and inverse approach that implies hypothetical models (not necessarily one) for the purpose of learning and/or deciding. The purpose of this paper is to elaborate on the fundamental elements that are contributing to the transformation of finance and raise its risk consequences.

1. INTRODUCTION

The growth of financial complexity, technology, computing capacities, and services combined with an access to "big" and varied data are currently transforming real finance and challenging its conventional models. Data science has for many generations challenged the practical implementation of theoretical models, statistical learning, numerical techniques, and their approximations. IT software and increasingly powerful computers have contributed to its new computing capacity and their ability to resolve conceived problems hitherto unapproachable. Concurrently, they have upended the search and usefulness of data algorithms based on statistical and computing facilities of various sorts. These are currently providing a range of opportunities to reduce costs and increase profits for financial institutions through the discovered potential of on-line systemic learning, trends discoveries, and their applications to become self-operated, often alluded to as AI (Artificial Intelligence) [Billard and Diday (2003), Callebaut (2012), Chambers (1993), Cleveland (2001), Donoho (2015), Hey et al. (2009), Kirkpatrick and Kurths (2012), National Research Council (2010)].

Developments, improvements, and the control of financial complexity are essential. For example, Ashby's [Ashby (1956)] Principle of Required Variety (The second Law of Cybernetics) already implied long ago that complexity untamed by mathematical intelligence and controls is self-destructive. In this vein, pursuing a financial evolution devoid of "intelligence" will necessarily lead to a financial and technological breakdown.

These approaches have been developed and used ever since the first Industrial Revolution, ushering in industrial automation, cybernetics, robotics, etc. Similarly,

mathematical algorithms have developed a multitude of algorithms to search, track, solve, and implement these same problems. Numerical optimization techniques, such as linear programming, stochastic modeling and optimization, and so on, have been devised to revolutionize our capacity to solve practical and complex economic, financial, industrial, service, design, and management problems. Current developments are adapted to a far broader set of applications such as self-driving cars, health care, finance, services, etc.

> "A machine learning algorithm walked into a bar. The bartender asked, "What would you like to drink?" The algorithm replied, "What's everyone else having?""
>
> — (SEEN ON TWITTER)

Models are tested and learning-improved by data, and inversely defined by data upended by learning processes based on both the information that data produces and statistical concepts that estimate and improve the belief in our models. This approach is trumpeted now as a means to confront and manage a complexity hitherto ignored. Terms, such as machine learning, deep learning, robo-advisors, artificial intelligence, and a multitude of algorithms are applied to specific problems deemed far too complex to be defined by just theoretical models. For some, it may seem that models defined ex-ante are "irrelevant" while "ex-post data" (since they are always ascertained "after the fact" or by simulations) is painted as a greater truth seizing the rationality that models portend to present. Namely, measurements that define events or conditions that express only "what is" rather than what we seek in the future to defines: "what to do." Data intelligence in such cases define both "what is" and "what to do" and the potential models (and thus rationalities) that underlie the data we dispose of. The complexity of "what is" and "what is to be" is far too great for one (models and statistics) not to be integrated without the other (the data approach). Models expressing strategic intents are then a "GPS," altered as new data is mined, analyzed, and applied to improve the "GPS" and edit the policies it implies. In this paper, we seek to appreciate ex-ante "modeling" and ex-post data management [Breiman (2001), Diday and Esposito (2003), Goodman and Wong (2009), Guetzkow (1959, 1962), Horton et al. (2015), Krohs and Callebaut (2007), Nyamabuu and Tapiero (2017), Tapiero (2013), Tapiero et al. (1975), Tukey (1962)].

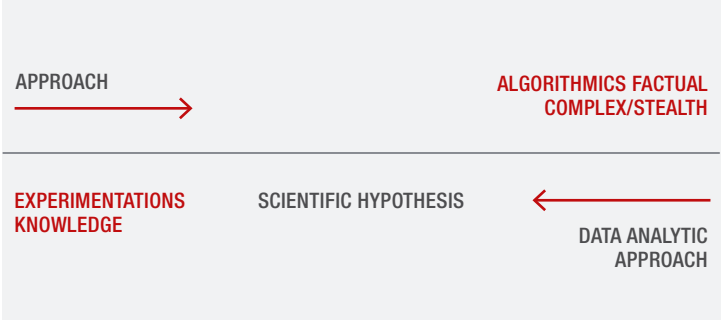
2. FINANCE AND DATA

The globalization of finance and financial technologies, combined with the complexities of financial systems and products led to a finance racing to transform its services, practices, trades, prices, and financial management, to be both data and computer operated and managed. For example, doing away with neighborhood banks and bank tellers, and replacing them with digitalized financial service and so on. It alters the role of traditional banks, financial markets, trade, insurance, etc. As information and technology become more accessible, new competitors are able to perform at relative and competing speeds and costs, increasing the efficiency of loan underwriting and credit scoring for individuals and SMEs. As a result, tasks that were predominantly performed by traditional banks are now computer aided with banks merely providing "financial" infrastructures for services such as trade and robotic advising and investments based on learning and inferential processes that might be designed by clouds of data and a software (algorithmic) intelligence [Albert and Barabási (2002), Overbye (2012), Nyambuu and Tapiero (2017)].

At the same time, banks have increased their push for a "cashless" future, setting the ground for a digitalized, technological, global, monopsonic, and competitive finance (although currently, faced with doubtful crypto-currencies challenges to financial regulation and money). Information Technologies (IT) and algorithms designed to meet the increased demands and the complexity of finance will, necessarily, face a global spiraling complexity challenging both strategic financial systems as well as their regulations. To confront this complexity, greater intelligence is required for both financial models and data. In such an environment, both creative and theoretical financial constructs, data mining and data analytics, statistical treatments that extract information, trends, strategies, decisions, and "models" may provide a finance architecture and the means to remain competitive, profitable, self-managed, and able to adapt to a future technological finance.

Practically, it means that finance is challenged by the intelligence that data provide and require. However, data without models seek a meaning to what they reveal by an inverse rationality. Namely, let data imply a model. Such an approach, necessarily, may provide not one, but many models. Explicitly, given the statistical character of financial data, it implies not one but many "models," maintaining the empirical presumption that all outcomes remain doubtful. For example, financial data prices are, by

Figure 1: Statistical versus data analytic approaches



definition, a present price, reflecting future potential prices and multiple factors such as yields, sentiments, news, macroeconomic trends, politics, the flow of domestic and foreign capital, etc. All of which affect a financial random future and expectations that define prices.

Options are such a case, with known parameters and prices implying a model of future volatility. Similarly, granular (fractional) financial data (reporting prices every day, every hour, second and microseconds) imply a financial randomness and information that data granularity provides [Tapiero and Vallois (2015, 2016, 2017, 2018a, b), Tapiero (2017)]. In such cases, data granularity defines the statistical properties it produces [Tapiero and Vallois (2015, 2016, 2017, 2018a, b), The Economist (2010), Vallois and Tapiero (2007, 2009)].

A complementary approach with one expanding the other [Allen (2001), Breiman (2001), Cleveland (1985, 1993), Donoho (2015), Groenen et al. (2006), Guttman (1944), Hanani and Tapiero (1980), Raveh and Tapiero (1980a, b), Tukey (1962, 1977, 1994)] leads to push-pull challenges that merge traditional ex-ante quants' complete markets models and a real and complex finance. These are, as stated above, defined by multiple factors such as macroeconomics, politics, globalization, complexity, emerging strategic (multi agent) finance, financial gating, and regulations (Nyambuu and Tapiero (2018), Raveh and Tapiero (1980b), Tapiero et al. (1975)].

3. DATA AND STATISTICAL LEARNING

Hardware and software have made it much easier to access mined data and transform it into information. Data access, storage, speed, and the growth of increasingly complex and integrated computer systems have in their wake opened new possibilities to "learn" using inferential software and generate future scenarios. Bayesian models, Copulas, long run (autocorrelation) memory, persistence (short memory), learning wavelets, Bayesian networks, neuro networks etc., provide predictive software, all of which are based on mathematical and statistical models. A comparison of data analytic approaches in Figure 1 emphasizes systemic causal approaches versus algorithmic-data analytic approaches.

These developments are "first generation" financial software combining computational and mathematical techniques with newly found computing and data management capabilities. Current intelligence software are fast mutating into future new generations that are hard to predict, however. For example, based on the presumption that data is never complete, information and the predictive powers it provides are also incomplete. The ancient Greeks, aware of the vagaries of time, claimed that "the likely is unlikely." These beliefs do not negate the fact that predictive models may be used. However, to mitigate their predictive uncertainty, a statistical rationale ought to be applied to qualify the quality of these predictions.

For example, inferences may provide a better appreciation of what data may teach us (revealed by errors and unconfirmed expectations we discover), as well as expand and qualify the breadth of our choices. Statistical learning can make predicting the future more efficient to the extent that the "machine" intelligence never outpaces financial human intelligence – an intelligence needed to be greater than the complexity it purports to confront. Lacking such an intelligence results in chaos [see also Ashby's Cybernetics (1956)].

Statistical models, by contrast to (ex-post) data as the sole mean to assess trends or inferences, are based on what we know in order to mitigate what we don't know applied to some specific and categorized purposes. Confirmed hypotheses are then used to predict future states, calculate and manage risks and their consequences, define "optimum" decisions and policies, and study their robustness. These in turn, produce a feedback mode that revises their hypothetical models.

Breiman (2001) pointed to two cultures. A learning culture that points to a model's (and statistical) uniqueness assessed by a model's statistical fit to improve and up-date, what we know, and mutate the model (hypotheses or estimated trends). By contrast, the data culture is an inverse approach that provides multiple potential "hypothetical models" that explain the data. Artificial intelligence is then a means to differentiate between these models, either explicitly or implicitly, to define a "common sense" and differentiate it from "nonsense" (based on a "cloud" of experiences and tested inferences).

Further, Breiman states that the roots of statistics, as in science, lie in working and checking theory against data. "I hope that in this (past) century our field will return to its roots (and that in the current century it will integrate data science and statistics)." There are today noticeable moves toward "real world data problems" and their integration with computer science, inverse theories, and to a greater awareness of consequences, social and otherwise.

Data analysis is traditionally associated with a statistical rationality based on learning, adaptive estimates, a long empirical and shared experience, and common theoretical tested principles [for example, see Andersen et al. (2009), Callebaut (2012), Goodman and Wong (2009), Hey et al. (2009), Krohs and Callebaut (2007), McKinsey (2011), Tukey (1962)]. For example, given sampled time series, models of stock prices are used to replicate, track, and infer future prices statistically qualified. Such approaches presume that data is incomplete, and therefore require mathematical and statistical principles to guarantee the quality of estimates and predictions. When data is "big" and presumed complete, it is fed by data analyses and algorithms. For example, Bayesian analysis provides a slow learning process based on pre-posterior estimates. In a changing and unstable environment, financial prices may define time series that are auto-correlated, with mutating trends, i.e., altering their statistical properties as events and time alter the model (providing, therefore, a learning process with very short time spans). Learning from data is, therefore, multi-dimensional, based on trends and the many facets and characteristics data implies. For example, rather than consider a time series trend, its mean and variance, additional factors such as the data samples' granularity, samples range, inverse statistics (of the time series surrogate processes) etc., provide additional dimensions along which data reveals its properties.

Similarly, and practically, data-memory is essentially an abstract filter that produces an image of the past based on data mined, transformed, and statistically treated (thereby, transforming the models that data implies). For example, option prices are implied by a specific "time and future limited horizon." Long run memories and their auto-covariance, co-location, and quantum entanglement altered events (producing a short term memory, with a stochastic mutation due to occurring events) are challenges that require an underlying model as well as a greater assessment of what data does reveal truly.

Breiman predicted these problems by pointing to three elements:

- **The Rashomon effect** (the Japanese film, same data, different perceptions): where a multiplicity of good models may result from the same data.
- **Occam razor:** relates to the conflict between simplicity and accuracy (for example: econometric models, and financial models in general, are not necessarily more accurate the greater their complexity).
- **Bellman:** the curse of dimensionality (and therefore, big data might not be more informative than "small" but "intelligent" data)

For example, Tukey (1962, 1977, 1994) predicted years ago that the future would emphasize the primacy of data [see also Donoho (2015)] and the need to learn from data analytics rather than just "fundamental statistical models." Tukey (1994) states that "For a long time I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt. ...I have come to feel that my central interest is in data analysis which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning and gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data."

Statistical and mathematical psychologists have been concerned with similar problems. Guttman (1944), for example, suggested a scaling approach to very large and multidimensional psychological (and educational) tests that led to the development of the "Guttman scale" ("a linearized multidimensional data). Sociology, psychology, and health data are often studied using very large quantities of qualitative data that include a large number of interactions, behavioral patterns, and variables. Each variable is defined in addition by its attributes. Scale can then be used to compare one student knowledge to another, a mental state compared to another by using Guttman's scaling methodology based on quantitative and qualitative data. Further development at the Bureau of Social Research in Jerusalem improved this approach by a dimensional reduction of multivariate datasets [see also Raveh and Tapiero (1980a, b)] and provided apparently more information than standard correlation and studies using factor analysis. Data analysis may then be easily assimilated in the form of "scalogram" providing a visual configuration of qualitative data. Louis Guttman's approach

was used successfully in investigating military morale and other problems (the U.S. Army Research Branch Morale Services Division) during WWII. Subsequently, it led to the development of numerous applications by the Bureau of Social Research (Jerusalem, Israel) on voting patterns and a broad range of questionnaires accumulated into large datasets.

Drawing on work of data scientists, a vision of data engineered to be better displayed rather than modeled have contributed also to better "learning from data." In other words, they have produced a friendlier data, visually accessible, providing an easier appreciation of what data means. For example, defining what a data user wants along any number of criteria and articulating a personalized vision of data. "Data science" was defined in terms of six divisions [Donoho (2015)]:

- Data exploration and preparation
- Data representation and transformation
- Computing with data
- Data modeling
- Data visualization and presentation
- Science about data science

These elements are parts and parcel of financial data science, applied to financial time series and other data types commonly used by financial agents. The first and second stages are crucial, especially when considering data preparation and transformation, as the success of some algorithms (such as deep learning and random forest algorithms that require some standardization of the data).

4. RISK MODELS AND DATA FINANCE

Financial theories derived from economic and risk models are hypotheses. Sometimes they are right, sometimes they are incomplete, always in doubt, and never confirmed. Financial risk models are merely partial models of uncertainty that predict future prices. Thus, risk management is based on predicting and accounting for the 'predictable,' rather than managing unpredictable and consequential risks, such as booms and busts, systemic risks, contagious behaviors, and so on, that recur mostly unpredictably. Risks, furthermore are consequences generated by multiple factors, from many sources, some of which are statistically and causally dependent. The following is a summary of some of its elements underlying data finance.

- Increased complexity and uncertainty and a belief in the certainty of data.
- Default models due to incomplete financial models.
- An increasingly strategic finance with dominant agents and a world at risk beset by what we do, by what others do, or we do to each other. In this world, a general equilibrium may no longer be possible, sustainable, or efficient. Risk finance is thus increasingly strategic.
- Financial greed, with TBTF (too big to fail) enterprises, and information and power asymmetries and increasingly aggressive and strategic regulations lead to the tenets of free markets to falter.
- Competing regulations in a global world contribute to increasingly complex financial logistics and to competing financial systems.
- Financial systems increasingly subdued to political and macroeconomic events are also faced with a far more complex risk finance, where risks are derivatives of non-finance risks.

These are partial processes changing tomorrow's world of finance. They may result in extreme behaviors fueled by excessive unfiltered information, far more apt at generating contagious behaviors and, therefore, 'financial runs.' Security risks, networks, and IT are also important candidates that redefine future financial risks. In this environment, conventional financial risk models are no longer relevant.

Though it must be said that financial risk products, such as insurance and lending, and risk markets, such as VIX and Carbone markets, are merely mechanisms for speculating and risk sharing (and an important part of financial activity). For example, credit risks and insurance coverage may be co-dependent, derived from a large number of transparent causes due to networked and IT systems and the availability of "big data" used to better assess borrowers' collaterals, risk history, wealth, etc. For regulators, big data provide a greater transparency of potential non-compliance to an increasingly complex regulatory table.

The promises and the risks of big data in finance are in their infancy. Some may allow the prospective integration of financial models with the many data clouds storing investors and personalized information; a power it can sell and provide to financial institutions; the risks to individual liberties and security clients assume, and so on. Data, thus, fuels a plethora of data analytic techniques to increase profits. An increasing number of software companies and start-ups are proposing 'black boxes' to interpret consumers' sentiments and intents using internet comments on stocks, financial assets, and variables

deemed pertinent to the financial environment. Algorithms and learning machines are then created to seek and interpret images to detect a 'flow of sentiments' that are claimed to be related to (and thereby be early predictors of) stock markets performance, and predict consumers' choices and their implied preferences. A rising tide of data driven algorithms is thus emerging and engulfing finance and business to become information and technologically dependent (and therefore, a growing source of risk).

These processes contribute to an extraordinary growth of information asymmetry risks and the misuse of information and insiders' trading risks. For example, say that a company hires a data scientist to determine the public's attitude towards that company and its CEO. It would be like paying a psychiatrist to hear what one wants to hear (since data analysis need not provide one set of conclusive observations). In big data, the chances of finding what one wants to hear and what may be a real fact are equally high. Searching for meaning in large datasets, without theories, may be like seeking the North Pole without a compass. For these and other reasons, big data based on the accumulation of private information is a growing source of risk that contributes to important security problems.

The traditional statistical approach, unlike data intensive treatments, is based on fundamental hypotheses to be refuted or not [Diday and Esposito (2003), Fisher (1936), Albert and Barabási, (2002), Billard and Diday (2003)]. Thus, the 'statistical/scientific' approach reveals 'uncertainty' and its risks from a given and tested knowledge base. Is an evolving process based on a cycle to hypothesize, measure, test, and confirm-or-not? The data driven approach, instead, is a statement of current facts, and a presumed certainty rather than recognizing that all knowledge is partial – embedded in a greater uncertainty that statistics qualifies.

For banks, traders, and suppliers of financial information and advice, data and information are becoming primary assets. The Economist reported that between 1990 and 2005, more than 1 billion people worldwide entered the middle class, and by 2013 the amount of data transferred over the internet will reach 667 exabytes annually. According to Cisco the quantity of data continues to grow faster than the ability of the network to carry. Companies like Amazon's Web Services, AT&T's Synaptic Hosting, AppNexus, GoGrid, Rackspace Cloud Hosting, the HP/ Yahoo/Intel Cloud Computing Test bed, the IBM/Google, and Micro Strategy BI Cloud, have provided various types of cloud services to ease these data storage problems (while at the same time setting data at risk). Currently, a variety of corporate clouds and data services have been commercialized, providing an increased access to data to an expanding (and competing) population of data managers and scientists for any purpose, including risk purposes.

The ability of data technologies (with social media having shown a way to handle and analyze vast amounts of unstructured data), to test our abilities to translate complex, diverse, and dynamic data sources into workable financial information remains unproven. At the same time, an expanding digitalized financial system is allowing context-specific analyses.

Information and/or knowledge extracted from digital records render financial banks' jobs easier, by diagnosing and detecting risks accurately, and assessing their clients' propensity to assume risks (and hence improve the overall profitability and "quality" of their services). Similarly, digitalized data may prevent cybercrimes more effectively and thus contribute to the increasingly complex systems of financial networks, e-financial markets, and an increasing financial retailing dependence. Despite the potential for big data and financial data analytics, it may hide risk consequences that have not been revealed. In terms of security, Michael de Crespigny, CEO at ISF, stated a few years ago that: "Only half of organizations surveyed by the ISF are using some form of analytics for fraud prevention, forensics and network traffic analysis, while less than 20% are using it to identify information related to subject matter requests, predict hardware failures, ensure data integrity or check data classification."

Few organizations recognized the benefits of information security, yet many were already using data analytics to support their core business. Currently, security is a prime concern, pointing to a "growing tree" of functions and technologies that renders "security" a dynamic "big business," big data and IT finance challenging. For example, the practical current mismatch of micro and macro financial market arbitrage seeking models – separating micro-economic considerations from macro ones and negating their underlying effects in micro (pricing) financial models. Such a mismatch leads to financial markets becoming 'incoherent.' For example, the mortgage-backed securities (MBS) crisis of 2008 was such a mismatch; combining the conditions of 'a home for everyone' based on low initiation costs and interests, with long run (and unsustainable) individual and systemic risks. Future models in finance may, therefore, be concerned fundamentally with risks they have not yet experienced.

Practically, data science is expanding digitalization, cloud computing, and computing enterprises. IT monopsonic and media sectors such as Google, Microsoft, Amazon, Apple, and a multitude of small firms have discovered that data is an asset that can be mined, sold, and resold. Artificial intelligence, created already in 1958, transformed mathematical models into intelligent software. One example is INRIA's expert optimal control system, self-designing a software used in Tapiero (1988). Software evolutions have also produced obsolescence risks. Cobol and Fortran, although used for some time, were replaced by a new "must" Lisp language and Lisp machines. They too became obsolete with C and C++, the new "kids in town." Today, we have R and Python that will probably mature and be replaced by other languages. Academic and research tendencies to emphasize empirical studies and data at the expense of mathematical models' integrity is also contributing to an additional brand of risk.

Nevertheless, financial IT and data science are providing immense opportunities that can also turn out to be an unwieldy process, victimized by the belief that a larger haystack may help to find a needle in that haystack. Yet, tamed big data can complement the statistical/scientific approach by providing an opportunity to reveal new hypotheses and new opportunities that can set such approaches on a more certain footing. Digitalized financial systems allow automatic context-specific interpretations, aggregation, and analysis of data (e.g., what information is relevant or not to a particular market or stock). For example, information and/or knowledge extracted from digital records can render financial bank jobs easier when diagnosing and detecting risky clients. Of course, it leads at the same time to the removal of human interventions from such processes. These opportunities have significant financial benefits but can also harbor social and financial risks, with society's risks enthralled in "artificially intelligent financial systems."

5. CONCLUSION: DATA SCIENCE AND STATISTICS CHALLENGED

Financial data science at its initial phase is expanding and challenging. The questions below summarize some of the issues it is challenged by:

- Is big data about looking for a needle in a haystack by adding hay?
- Is the future of finance a data science without "models"?
- Is data science merely another IT data-driven tool? Compared to statistics that seeks to justify what we define or conclude based on data? Can data science decide what we are to do? Or merely advise and maintain the freedom of choice?
- Is big data and its businesses the end of privacy?
- Are algorithmic models processing data science transparent models? Are they means or ends? Do they reveal the unexpected or merely the expected?
- Is data science something new? Or the marketing of well-known data analytics tools up-ended with a greater computational efficiency (computers on steroids)?
- Is big data a means to increase or reduce complexity? If so, what are its consequences to regulation, compliance, and safe finance? Is the growth of complexity designed and part of data science?
- Is big data in finance an evolving artificial intelligence for the "war of machines"?
- Is the growth of data and its practical analysis sustainable?

By contrast, there are already payoffs to financial technology and data science including among others:

- Strategic positioning in a global increasingly monopsonic and competing gated world.
- The mass customization of products and services (and their yet undefined consequences).
- Does globalization, complexity, gating, and the transformation of increasing speed of financial and technological finance render finance one of politics and gated national policies or one of markets?
- Immediacy: the need to be here and there and trade everywhere and at all times as well evading regulatory regimes. Do these empower financial corporate firms at the expense on financial investors or vice versa.
- The need to communicate and to sustain a state of instant and mobile communication.
- The need to keep pace and paces away from a future falling upon us faster and faster than the present can handle. Would the relative adaptation of individual investors overwhelm the architecture of corporate financial and banking systems?

These are engines motivating technological and financial growth, and in their wake, the growth of IT networks and services launched daily by the Internet and IT enterprises. In this process, a financial IT infrastructure is:

- Growing ever more complex.
- More diffused, technologically and otherwise.
- Harder to define.
- Network based.
- More difficult to assume and control.

The achievements of industries in integrating IT and data are, by comparison to financial and other services, immense. For example, in industry Internet has contributed

to reductions in cycle times, costs, and labor, and helped introduce multiple products, and improve their qualities. It has also maximized and facilitated contacts across national boundaries leading to the globalization of industries and the expansion of supply networks. For these reasons, they incentivized services and financial technologies. Financial enterprises are proceeding in similar directions by adapting data science to theirs needs and expanding into a digitalized finance with networks and services spanning the globe. They do so to reduce financial logistic costs, expand their outreach, and increase their profits and the dominance of finance. Network effects have amplified innovation and creativity, and, in some cases, have empowered clients (although they have also augmented their hold of depositors). By the same token, it has expanded financial services to a far broader set of financial investors and empowered clients to manage their wealth globally. It has integrated financial and data systems and accessed online transactions, support multiple payment methods, and altered the business of banking.

REFERENCES

Albert, R., and Barabási A-L., 2002, "Statistical mechanics of complex networks," Reviews of Modern Physics 74:1, 47-97

Allen, J.F., 2001, "Hypothesis, induction and background knowledge. Data do not speak for themselves," BioEssays 23, 861–862. 33

Andersen T. G., R. A. Davis, J. P. Kreiss, T. Mikosch (eds.), 2009, Handbook of financial time series, Springer Verlag

Ashby W. R., 1956. Introduction to Cybernetics, Chapman & Hall

Billard, L., and E. Diday, 2003, "From the statistics of data to the statistics of knowledge: symbolic data analysis," Journal of the American Statistical Association 98:462, 470-487

Breiman, L., 2001, "Statistical modeling: the two cultures," Statistical Science 16:3, 199–231

Callebaut, W., 2012, "Scientific perspectivism: a philosopher of science's response to the challenge of big data biology," Studies in History and Philosophy of Biological and Biomedical Sciences 43:1, 69–80

Chambers, J. M., 1993, "Greater or lesser statistics: a choice for future research," Statistics and Computing, 3:4, 182-184

Cleveland, W. S., 2001, Data Science: an action plan for expanding the technical areas of the field of statistics," International statistical review, 69:1, 21-26

Cleveland, W. S., 1993, Visualizing data, Hobart Press

Cleveland, W. S. 1985, The elements of graphing data, Wadsworth Advanced Books and Software Monterey

Diday E., and F. Esposito, 2003, "An introduction to Symbolic Data Analysis and the Sodas Software," International Journal on Intelligent Data Analysis 7:6, 583–601

Donoho, D., 2015, 50 years of Data Science, September 18, version 1.00 (on the internet)

Fisher R. A., 1936, "The use of multiple measurements in taxonomic problems," Annals of Eugenics 7:2, 179-188

Goodman A. A., and C. Wong, 2009, "Bringing the night sky closer: discoveries in the data deluge," in Hey, T., The fourth paradigm: data-intensive scientific discovery, Microsoft Research

Groenen, P. J. F., S. Winsberg, O. Rodriguez, and E. Diday, 2006, "I-Scal Multidimensinal scaling of interval dissimilarities," Computational Statistics and Data Analysis 51, 360-378

Guetzkow, H., 1959, "A use of simulation in the study of inter-nation relations," Behavioral Science 4:3, 183-191

Guetzkow, H., 1962, Simulation in social science: readings, Prentice-Hall

Guttman L., 1944, "A basis for scaling qualitative data," American Sociological Review 9:2, 139-150

Hanani, U., and C. S. Tapiero, 1980, "The combined location – districting problem: a dynamic clusters analysis approach," Analyse de Donnees, INRIA, 249-266

Hey, T., S. Tansley, and K. Tolle, (eds.), 2009, The Fourth Paradigm: data-intensive scientific discovery, Microsoft Research

Horton , N. J, B. S. Baumer, and H. Wickham, 2015, "Setting the stage for data science: integration of data management skills in introductory and second courses in statistics," arXiv preprint arXiv:1502.00318

Kirkpatrick, S., and J. Kurths, 2012, "Challenges in network science: applications to infrastructures, climate, social systems and economics," European Physics Journal Special Topics 214:1, 273-293

Krohs, U., and W. Callebaut, 2007, "Data without models merging with models without data," in Boogerd, F. C., F. J. Bruggeman, J. H. S. Hofmeyr, and H. V. Westerhoff (eds.), Systems biology: philosophical foundations, Elsevier

Knight, F. H., 1921, Risk, uncertainty and profit, Cornell University Library

Langlois, R. N., and M. M. Cosgel, 1993, "Frank Knight on risk, uncertainty, and the firm: a new interpretation," Economic Inquiry 31:3, 456–465

McKinsey, 2011, "Big data: The next frontier for innovation, competition, and productivity," https://mck.co/2dDfLuS

National Research Council, 2010, "Technical capabilities necessary for regulation of systemic financial risk: summary of a workshop," The National Academies Press, Washington, D.C.

Nyambuu, U. and C. S. Tapiero, 2018, Globalization, gating and risk finance, Wiley

Pearl, J., 1985, "Bayesian networks: a model of self-activated memory for evidential reasoning," (UCLA Technical Report CSD-850017). Proceedings of the 7th Conference of the Cognitive Science Society, University of California, Irvine

Raveh, A., and C. S. Tapiero, 1980a, "Periodicity, constancy, heterogeneity and the categories of qualitative time series," Ecology 61:3, 715-719

Raveh, A., and C. S. Tapiero, 1980b, "Finding common seasonal patterns among time series: an MDS approach," Journal of Econometrics 12, 353-363

Tapiero, C. S., 2013, Engineering risk and finance, Springer

Tapiero, C., S., 1988, Applied stochastic models and control in management, North-Holland

Tapiero, C. S., M. F. Capobianco, and A. Y. Lewin, 1975, "Structural inference in organizations," Journal of Mathematical Sociology 4, 121-130

Tapiero, C. S., and P. Vallois, 2000, "The inter-event range process and testing for chaos in time series," Neural Network World 10:1–2, 89–99

Tapiero, C. S., and P. Vallois, 2015, "Financial modelling and memory: Mathematical system. Future perspectives in risk models and finance," Springer, Vol. 211, 149-246 (International Series in Operations Research and Management Science; vol. 211)

Tapiero, C. S., 2017, "Data science and intelligence," Risk and Decision Analysis 6:4, 291-298

Tapiero, C. S., and P. Vallois, 2016, "Fractional randomness," Physica A: Statistical Mechanics and its Applications 462, 1161-1177

Tapiero, C. S., and P. Vallois, 2017, Implied fractional rates and default risk distributions, Probability, Uncertainty and Quantitative Risk, Springer.

Tapiero, C. S., and P. Vallois, 2018a, "Fractional randomness and the Brownian bridge," Physica A 503, 835–843

Tapiero, C. S., and P. Vallois, 2018b, "Fractional randomness and alpha-stable distributions," Physica A 511, 54-60

Overbye, D., 2012, "The trouble with data that outpaces a theory," March 26, New York Times, https://nyti.ms/2wTt0AT

The Economist, 2010, "Data, data everywhere," February 25, https://econ.st/2x1TNdw

Tukey, J. W., 1962, "The future of data analysis," The Annals of Mathematical Statistics 33:1, 1-67

Tukey, J. W., 1977, Exploratory data analysis, Pearson

Tukey, J. W., 1994, The collected works of John W. Tukey, Vol. 1, Taylor & Francis

Vallois, P., and C. S. Tapiero, 2007, "Memory-based persistence in a counting random walk process," Physica A 386:1, 303-317

Vallois, P., and C. S. Tapiero, 2009, "A claims persistence process and insurance," Insurance Economics and Mathematics 44:3, 367-373

SUPERVISION

EARLY WARNING INDICATORS OF BANKING CRISES: EXPANDING THE FAMILY[1]

IÑAKI ALDASORO | Economist, Monetary and Economic Department, BIS
CLAUDIO BORIO | Head of the Monetary and Economic Department, BIS
MATHIAS DREHMANN | Principal Economist, Monetary and Economic Department, BIS

ABSTRACT

Household and international debt (cross-border or in foreign currency) are a potential source of vulnerabilities that could eventually lead to banking crises. We explore this issue formally by assessing the performance of these debt categories as early warning indicators (EWIs) for systemic banking crises. We find that they do contain useful information. In fact, over the more recent subsample, for household and cross-border debt indicators the information is similar to that of the more commonly used aggregate credit variables regularly monitored by the Bank for International Settlements (BIS). Confirming previous work, combining these indicators with property prices improves performance. An analysis of current global conditions based on this richer information set points to the build-up of vulnerabilities in several countries.

1. INTRODUCTION

Early warning indicators (EWIs) of banking crises are typically based on the notion that crises take root in disruptive financial cycles. The basic intuition is that outsize financial booms can generate the conditions for future banking distress. The narrative of financial booms is well understood: risk appetite is high, asset prices soar, and credit surges. Yet, it is difficult to detect the build-up of financial booms in real time and with reasonable confidence. It is here that EWIs come in. Many studies, including at the BIS, have found that one can identify such unsustainable booms reasonably well based on, say, deviations of credit and asset prices from long-run trends (gaps) breaching certain critical thresholds.

To detect the build-up of vulnerabilities around the globe, in recent years the BIS has regularly published credit-to-GDP gaps, economy-wide debt service ratios (DSRs), and, less frequently, property price gaps. While these aggregate indicators are useful, research has pointed to the importance of specific subcategories of debt as a source of vulnerabilities, especially household debt and cross-border and/or foreign currency debt (international debt for short).

Against this backdrop, in this article we do two things. First, we assess the EWI performance of both household and international debt. Here, we draw extensively on BIS statistics, notably the international banking and securities statistics as well as sectoral credit data and DSRs. Second, we map the statistical analysis into current conditions, taking into account also the information from aggregate indicators. In the process, we take the opportunity to clarify some common misconceptions about the interpretation of EWIs.

[1] The authors would like to thank Stefan Avdjiev, Stijn Claessens, Ben Cohen, Ingo Fender, Mikael Juselius, and Pat McGuire for helpful comments and Bat-el Berger, Anamaria Illes, Matthias Lörch, Kristina Micic, and Taejin Park for excellent research assistance. The views expressed in this article are those of the authors and do not necessarily reflect those of the BIS.

We come up with three main findings. First, indicators based on household and international, in particular cross-border, debt do contain useful information about future banking distress. The household sector DSRs perform especially well. Second, as might be expected, it is possible to further improve performance by combining individual indicators. Confirming previous work, we find that combining debt variables with property prices is especially helpful. Finally, the indicators currently point to the build-up of risks in several economies.

This article is structured as follows. The first section reviews the rationale behind EWIs and provides a first look at the behavior of selected candidate indicator variables around crises. The second formally evaluates their EWI performance. We first compare indicators on a standalone basis and then show the gains in predictive power from combining them. The third section draws on these findings to discuss current vulnerabilities. We also provide a short guide on how to interpret indicators. A final section concludes.

2. THE RATIONALE BEHIND EWIs

EWIs typically capture booms in the financial cycle in a stylized way. The notion of the financial cycle refers to the self-perpetuating sequence of financial expansions and contractions that can amplify business fluctuations [Minsky (1982), Kindleberger (2000), Borio (2014)]. Further, outsized financial booms can lead to stress and even financial crises. While progress has been made, measuring the financial cycle remains challenging: the underlying theoretical modeling is scant and there is no single aggregate measure of financial activity. That said, a consensus has started to emerge that credit aggregates and asset prices, especially property prices, play an important role [Terrones et al. (2011), Drehmann et al. (2012)].

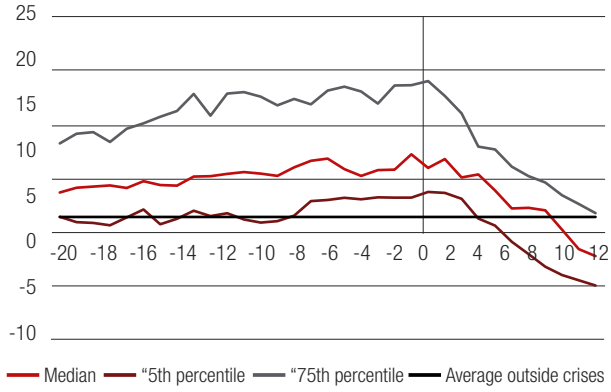
The existing BIS EWIs translate the intuitive notion of a financial boom into simple and transparent measures. The BIS has regularly published and monitored aggregate private sector credit-to-GDP gaps, residential property price gaps, and DSRs for the private non-financial sector. The credit-to-GDP gap is calculated as the difference between the credit-to-GDP ratio and its (one-sided) long-term trend.[2] Detrending is designed to remove the impact of benign, long-term changes in the underlying series; for example, those that result from financial development. The gap opens up if the increase in the credit-to-GDP ratio strongly outpaces the trend for some time, pointing to a possible financial imbalance. The property price gap is the equivalent measure, defined as the deviation of inflation-adjusted property prices from their trend. DSRs measure interest payments and amortizations relative to income.[3] As high credit growth feeds into higher debt service down the road, DSRs rise during credit booms [Drehmann et al. (2017)]. And since they take into account interest payments, they could perform better than the credit gap or credit growth when debt builds up continuously but more slowly over time, making balance sheets vulnerable to increases in interest rates.

[2] The credit-to-GDP gap is the difference between the ratio of total non-financial sector credit to GDP and its trend based on a one-sided Hodrick-Prescott (HP) filter with the smoothing parameter equal to 400,000. Such a high value ensures a very slowly moving trend. The residential property price gap is the deviation of inflation-adjusted residential property prices from a similarly constructed trend. For a discussion of the appropriateness of this trend measure in this specific context, see Drehmann and Tsatsaronis (2014).

[3] Since most countries do not compile data on amortization payments, these are estimated using information from debt maturities, interest rates, and outstanding debt stocks [Drehmann et al. (2015)].

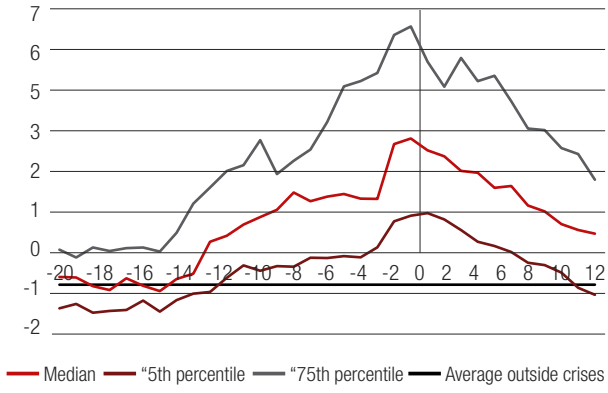
Figure 1: Evolution of existing BIS EWI variables around past banking crises

CREDIT-TO-GDP GAP



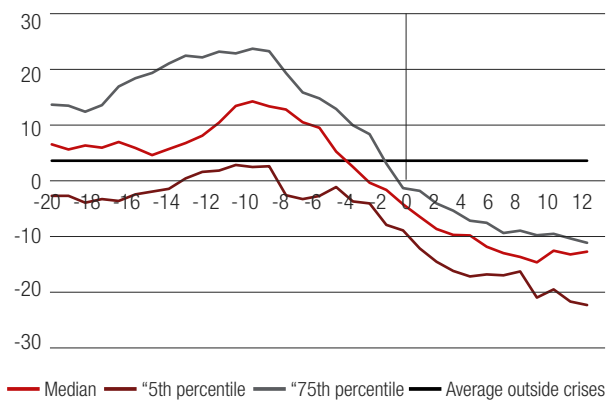
Difference of the credit-to-GDP ratio from its long-run trend computed with a one-sided HP filter.

TOTAL DSR



Difference of the total DSR from country-specific 20-year rolling averages.

PROPERTY PRICE GAP



Deviation of real property prices from their long-run trend computed with a one-sided HP filter.

Note: The vertical line indicates time = 0. The historical dispersion (median, 25th, and 75th percentiles) of the relevant variable is taken at the specific quarter across all crisis episodes available for the respective indicator.

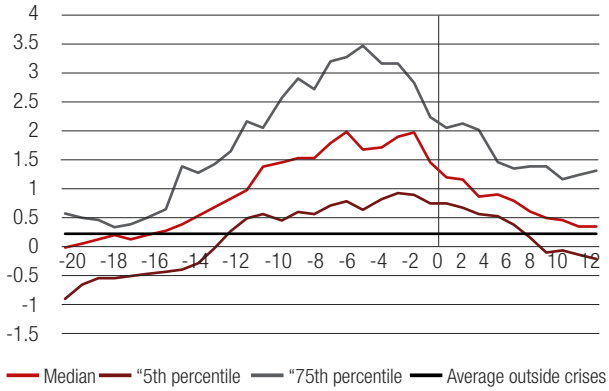
Sources: IMF, International Financial Statistics; national data; BIS credit to the non-financial sector and property price statistics; authors' calculations.

It is thus unsurprising that credit, DSRs, and property price gaps were comparatively high before past crises (Figure 1). For much the same reason, they perform well as EWIs on a standalone basis, and even better if combined [Borio and Lowe (2002a), Drehmann et al. (2011), Drehmann and Juselius (2012), Detken et al. (2014)].[4]

In addition to the aggregate credit developments covered by the current BIS EWIs, recent research has highlighted the importance of the household sector specifically. While higher household debt boosts consumption and output growth in the short run, too much of it can lower output growth in the medium to long term [Mian et al. (2017), Lombardi et al. (2017), Zabai (2017)]. Excessive household debt has also been found to herald banking crises [Jordà et al. (2016), IMF (2017), Drehmann et al. (2017)]. As such, indicators assessing household debt developments feature prominently in many central bank financial stability reports [Bank of Canada (2017), ECB (2017), Bank of England (2017)].

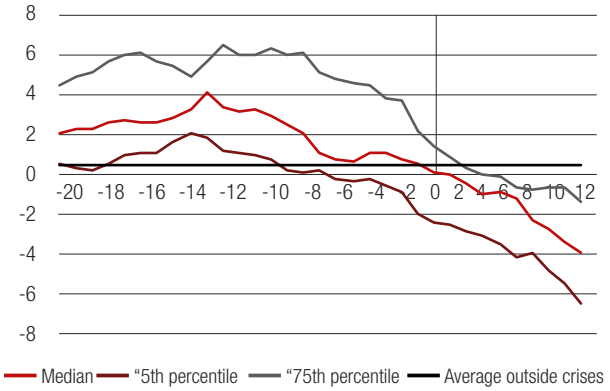
Figure 2: Evolution of new EWI variables around past banking crises

HOUSEHOLD DSR



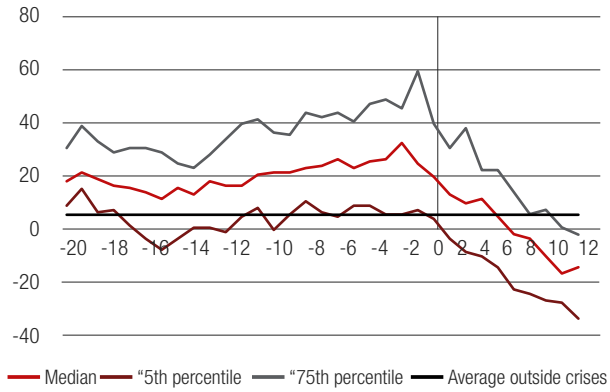
Difference of the household DSR from country-specific 20-year rolling averages.

HOUSEHOLD CREDIT-TO-GDP GAP



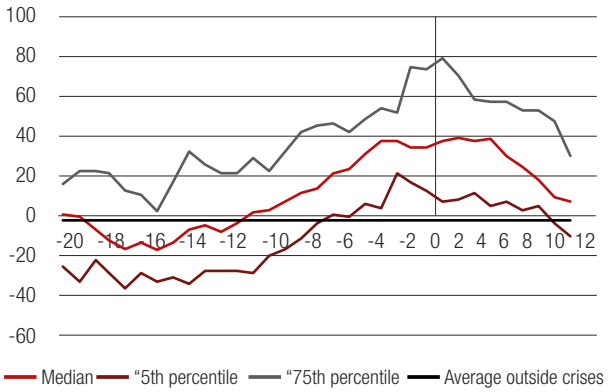
Difference of the household credit-to-GDP ratio from its long-run trend computed with a one-sided HP filter.

CROSS-BORDER CLAIMS TO GDP



Twelve-quarter growth rate in the cross-border claims-to-GDP ratio.

FOREIGN CURRENCY DEBT TO GDP



Twelve-quarter growth rate in the foreign currency debt-to-GDP ratio.

Sources: IMF, International Financial Statistics; national data; BIS credit to the non-financial sector, debt securities, locational banking and property price statistics; authors' calculations.

Note: The vertical line indicates time = 0. The historical dispersion (median, 25th, and 75th, percentiles) of the relevant variable is taken at the specific quarter across all crisis episodes available for the respective indicator.

[4] The credit gap was first proposed by Borio and Lowe (2002a), and the literature has found broadly similar EWI performance for slightly different measures, such as five-year growth rates in the credit-to-GDP ratio [Schularick and Taylor (2012)]. The credit-to-GDP gap has been incorporated into the policy process as the trigger variable for the imposition of a countercyclical capital buffer on supervised banks [BCBS (2010)].

We consider two household sector indicators. The first is the household credit-to-GDP gap – an exact analogue of the total credit-to-GDP gap but using only credit to households in the numerator.[5] The second is the difference between the household sector DSR and its 20-year rolling average [Drehmann et al. (2017)].[6] By normalizing with a one-sided trend or a rolling average, we try to mimic the real-time environment policymakers face: the indicators are only based on past information, available at the time decisions are made.

Policymakers have also long focused on foreign currency and/or cross-border debt as a source of financial stability risks [Bruno and Shin (2015), Chui et al. (2014), BIS (2017), Borio et al. (2011), Avdjiev et al. (2012)]. In part, because of data limitations, the EWI literature has operationalized this by looking at current account deficits [Lo Duca and Peltonen (2013)] or exchange rate developments [Borio and Lowe (2002b) and Gourinchas and Obstfeld (2012)].[7]

Drawing on the BIS international banking and debt statistics, we go one step further and explicitly evaluate cross-border borrowing, as well as foreign currency debt, issued across borders and at home. To normalize by country size and to tease out medium-term developments, we take the three-year growth rates in the corresponding ratios to GDP.[8] The foreign currency debt is that of non-banks. For cross-border claims, we take a broader perspective that captures lending to non-banks and banks.[9] We do so as indirect cross-border credit, i.e., cross-border credit that banks lend on to non-banks, is a frequent enabler of domestic credit expansions [Avdjiev et al. (2012)].[10]

Data coverage differs across indicators.[11] We have credit-to-GDP gaps and cross-border credit for 42 jurisdictions, often from the first quarter of 1980 to the second quarter of 2017.[12] Data are most limited for the household DSR, which is only available for 27 jurisdictions and often starts only in the mid-1990s. For crisis dating, we rely on the new European Systemic Risk Board crisis dataset [Lo Duca et al. (2017)] for European countries and on Drehmann et al. (2010) for the rest.[13]

A first glance at the data indicates that household debt may provide useful signals of the build-up of vulnerabilities (Figure 2). The household sector DSR (top row, left-hand panel) has been unusually high in the run-up to crises. The household credit-to-GDP gap (top row, right-hand panel) has also tended to be above normal levels during those phases.

The same holds for the international debt indicators (Figure 2, bottom row). The growth rate of the foreign currency debt-to-GDP ratio increases strongly pre-crisis, though it exhibits relatively high variation across countries (dashed lines). That of the cross-border debt-to-GDP ratio is also markedly higher but less variable.

[5] We also assessed the three- or five-year growth rate of the household credit-to-GDP ratio. This did not have a statistically significantly different performance from the household credit-to-GDP gap.

[6] As there are country-specific differences in the level, it is important to remove the long-run trend [Drehmann et al. (2015)].

[7] We also considered exchange rates and current account balances as indicators. But as they underperformed cross-border credit indicators, we exclude them from the reported results.

[8] Foreign currency debt is composed of the sum of US dollar-, euro-, yen-, sterling- and Swiss franc-denominated debt in the form of cross-border loans to non-banks, international debt securities issued by non-banks, and, where reported, local loans in foreign currency to non-banks. The series start in 1995, and we extend them backwards by applying the change in cross-border claims on non-banks from the BIS locational banking statistics. Our indicator on cross-border claims comprises lending in all instruments and currencies, to both banks and non-banks, as reported in the locational banking statistics. For both series we take the stocks and adjust them for breaks due to methodological or coverage changes. Given large breaks prior to 1984, we start from that point. (Available in the Online Appendix, obtainable from the authors).

[9] In addition to the growth rate in the gross claims relative to GDP, we also assessed the performance of a corresponding net indicator (claims minus liabilities). This is likely to be a better measure of the credit that remains within the country. That said, this variable did not perform as well as its gross counterpart.

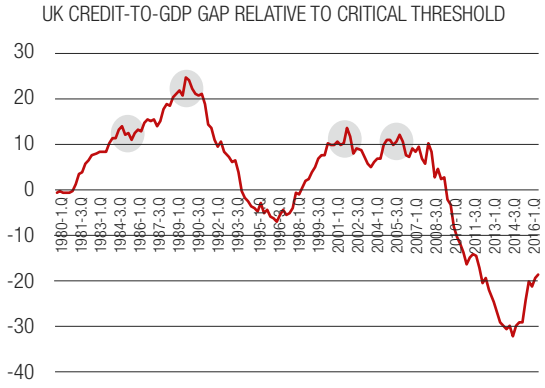
[10] Indirect credit is not included in the foreign currency debt series as we would run into problems of double-counting. For instance, a bank may borrow in foreign currency from abroad to lend domestically (also in foreign currency).

[11] Coverage and sources are discussed in detail in the Online Appendix, available from the authors.

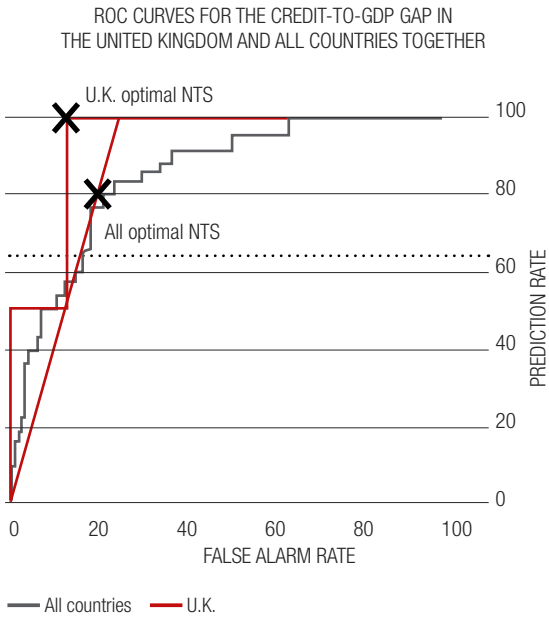
[12] Our broadest sample includes Argentina, Austria, Australia, Belgium, Brazil, Canada, Chile, China, Colombia, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hong Kong SAR, Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Korea, Malaysia, Mexico, the Netherlands, Norway, New Zealand, Poland, Portugal, Russia, Saudi Arabia, Singapore, Spain, South Africa, Sweden, Switzerland, Thailand, Turkey, the United Kingdom, and the United States.

[13] We exclude crises related to transitioning economies or that were imported from abroad based on Lo Duca et al. (2017). In addition, we classify the crisis in 2008 in Switzerland as imported. For the statistical analysis we drop post-crisis periods as identified in Lo Duca et al. (2017) and Laeven and Valencia (2012) for non-European countries.

Figure 3: Correct calls, false alarms, and the mapping between ROC curves and thresholds



The horizontal line at value 9 indicates the critical threshold for the credit-to-GDP gap obtained in Table 2. The shaded periods indicate the 12 quarters prior to crises – the prediction horizon. The black circles indicate periods in which the credit-to-GDP gap exceeded the critical threshold yet no crisis materialized within the prediction horizon. The red circles indicate periods in which the credit-to-GDP gap exceeded the critical threshold during the prediction horizon.



The horizontal line indicates a crisis prediction rate of 66%. The black crosses show the points on the ROC curves with the optimal noise-to-signal ratios. The green dotted line is the steepest line from the origin that touches a corner point on the portion of the ROC curve that is at or above the 66% prediction rate, identifying the point with the lowest noise-to-signal ratios.

Sources: IMF, International Financial Statistics; national data; BIS credit to the non-financial sector statistics; authors' calculations.

3. EVALUATING EWIs

When formally evaluating the performance of the EWIs, one would ideally like to know how policymakers assess the trade-off between missed crisis calls (type I errors) and false alarms (type II errors). However, this cannot be done with any precision, not least due to the limited experience from which to estimate expected costs and benefits [CGFS (2012)].

Absent well specified trade-offs, one way to assess the performance of EWIs is to consider the full mapping between type I and type II errors. This mapping is called the receiver operating characteristic (ROC) curve (see Section 3.1 for details). The area under the curve (AUC) is a convenient and interpretable summary measure of the signaling quality of a binary (yes/no) signal. A completely uninformative indicator has an AUC of 0.5. Correspondingly, the AUC for the perfect indicator equals 1. The AUC of an informative indicator falls in between and is statistically different from 0.5.

The AUC is a useful starting point, but it does not provide any information about the critical thresholds that, if breached, should raise concerns about financial stability risks. These ultimately depend on policymakers' preferences. To derive the thresholds, we assume that policymakers choose one that minimizes the noise-to-signal ratio (the ratio of false alarms to correctly predicted events) while capturing at least two-thirds of the crises, as in Borio and Drehmann (2009). (The section below discusses the link between this criterion and the ROC curve).

To be useful for policy, EWIs should not only have statistical forecasting power and rely on real-time information, but also satisfy three additional requirements [Drehmann and Juselius (2014)]: timing, stability, and ease of interpretation.

Having the right timing means that the indicators' signals should arrive early enough so that policy measures can be implemented and have an impact. That said, signals that arrive too early can be problematic [Caruana (2010)]. We focus on a 12-quarter forecast horizon.[14] Employing a multi-year horizon also recognizes that the indicators may help identify the build-up of vulnerabilities, but cannot be expected to pinpoint the specific timing of a crisis.

[14] Strictly speaking, one could drop the year that precedes the crisis on the grounds that by then it would be too late to take major preventive steps.

EWIs should also provide stable signals. Policymakers prefer to react to persistent movements, given the uncertainties involved. Stability requires that the forecast performance should not decrease as crises approach. This is a problem for residential property prices [Drehmann and Juselius (2012)], for which growth tends to slow or even become negative closer to crises (Figure 1, left-hand panel). This makes it hard to discern in real time whether the slowdown reflects the typical pre-crisis behavior of property prices or a welcome correction.

Finally, unless EWIs are easy to interpret intuitively their signals are likely to be ignored [Önkal et al. (2002), Lawrence et al. (2006)]. This is why our EWIs are simple, transparent, and based on the financial cycle logic. Their simple structure may also reduce the risk of overfitting associated with more sophisticated techniques.

3.1 Evaluating EWIs: ROC curves, noise-to-signal ratios and critical thresholds

It is possible to illustrate the trade-off between correct event predictions (as a share of all events) and false alarms (as a share of all normal periods) when choosing a threshold in the case of the credit-to-GDP gap for the United Kingdom (U.K.). The left-hand panel of Graph A shows the evolution of the gap since 1980. The shaded areas highlight the three years before the crises in 1991 and 2007, the period when we would like to see a signal based on the assumed three-year prediction horizon. The dashed red horizontal line indicates a credit-to-GDP gap of 9 – the optimal threshold given our analysis (Table 2). In both pre-crisis periods, the gap exceeded 9, so the prediction rate is 100% (red circles). Yet there are also false alarms (black circles). Increasing the threshold above 9 reduces the number of false calls. But once the threshold exceeds 11.5, the crisis in 2007 is no longer predicted, so that the prediction rate falls to 50%. Conversely, lowering the threshold from 9 does not increase the prediction rate and leads only to more false alarms.

The receiver operating characteristic (ROC) curve captures this trade-off between correct predictions and false alarms for all thresholds. For the U.K. the prediction rate can only be 100%, 50%, and 0% (Figure 3, top panel, blue line), with false alarm rates decreasing as the threshold increases. The solid red line depicts the ROC curve for the credit-to-GDP gap based on all the available data in our sample. We can see that the credit-to-GDP gap is an informative indicator but is not perfect. For a perfect indicator we would find at least one threshold with a prediction rate of 100% and a false alarm rate of 0%. At the other end of the spectrum, a completely uninformative indicator would have an ROC curve that equaled the 45° line for every threshold, i.e., the same rate of correct and false calls.

The area under the ROC curve (AUC) provides a summary measure of the signaling quality of an indicator. Intuitively, it captures the average gain over the uninformed case (the 45° line) across all possible threshold combinations. The uninformative indicator has an AUC equal to 0.5 (i.e., the area under the 45° line equals 0.5), while that for the fully informative indicator is equal to 1. The intermediate cases have values in between.

While the ROC maps the full set of trade-offs, the policymaker may weigh missed crisis calls and false alarms differently. Unfortunately, these preferences are not known. As discussed in the main text, we therefore assume that policymakers choose a threshold that minimizes the noise-to-signal ratio (the ratio of false alarms to correctly predicted events), while capturing at least two-thirds of the crises.

It is possible to find the points on the ROC curve that correspond to the optimal thresholds for the U.K. and the more general case (black crosses). The U.K. case is especially intuitive. One picks the part of the ROC curve that identifies a prediction rate of at least 66% of crises – here the only possible one is 100%. Next one moves on that line as far as left as possible, thereby minimizing false alarms, i.e., one chooses the leftmost corner. The more general case is slightly more complicated, although the procedure is the same. One picks the steepest line from the origin (dotted green line) that touches a corner point on the portion of the ROC curve that is at or above the 66% prediction rate (red dashed line). This works because the slope of such a line equals the signal-to-noise ratio of the threshold associated with the corner point on the ROC curve. And as the signal-to-noise ratio is the inverse of the noise-to-signal ratio, the steepest line finds the point on the ROC curve with the lowest noise-to-signal ratio.

Table 1: A comparison of the predictive power of single EWIs using the AUC

EWI	HORIZON (QUARTERS)											
	1	2	3	4	5	6	7	8	9	10	11	12
CREDIT-TO-GDP GAP	0.80*	0.80*	0.81*	**0.80***	**0.79***	**0.78***	**0.75***	**0.78***	**0.77***	**0.78***	**0.76***	**0.77***
TOTAL DSR	**0.84***	**0.83***	**0.82***	0.79*	0.77*	0.76*	0.74*	0.75*	0.74*	0.71*	0.69*	0.69*
PROPERTY PRICE	0.43	0.49	0.54	0.61	0.64	0.67*	0.68*	0.72*	0.71*	0.70*	0.69*	0.68*
HOUSEHOLD DSR	0.82*	0.81*	0.80*	0.79*	0.77*	0.76*	0.73*	0.76*	0.75*	0.72*	0.69*	0.67*
HOUSEHOLD CREDIT-TO-GDP GAP	0.60	0.60	0.61	0.65*	0.66*	0.67*	0.70*	0.74*	0.75*	0.76*	0.76*	0.76*
FOREIGN CURRENCY DEBT TO GDP	0.73*	0.71*	0.71*	0.67*	0.64	0.60	0.52	0.47	0.49	0.46	0.46	0.45
CROSS-BORDER CLAIMS TO GDP	0.75*	0.75*	0.77*	0.74*	0.73*	0.71*	0.67*	0.66*	0.65*	0.63*	0.60	0.58

An asterisk denotes that the estimated AUC is statistically significantly higher than 0.5. The numbers in bold indicate the highest AUC for each horizon. Numbers in green indicate that the AUC is not statistically significantly different from the highest one at this horizon.

Sources: IMF, International Financial Statistics; national data; BIS credit to the non-financial sector, debt securities, locational banking and property price statistics; authors' calculations.

4. STANDALONE INDICATORS

To evaluate and compare the performance of the indicators on a standalone basis, we proceed in two steps. Initially, to assess their general information content, we use the AUC criterion. We then evaluate the indicators from an operational perspective by analyzing optimal thresholds based on specific preferences.

We do so using two different samples: the full sample available for each indicator, and the much smaller common sample. The common sample allows a comparison of like with like, but it reduces our sample size considerably. We, therefore, also use the full sample available for each indicator as a comparison.

Although we try to collect as much data as possible, predicting crises inevitably means predicting rare events. Data coverage is best for the credit-to-GDP gap. But even then, we only cover 30 crises. The common sample covers 19 episodes, 12 of which are related to the Great Financial Crisis (GFC). In addition, the dataset is tilted towards advanced economies. Thus, the use of the full sample available for each indicator is important for robustness. For brevity, we only report this for the threshold analysis. In addition, we did robustness checks, not reported here, running the statistical tests on pre- and post-2000 subsamples to ensure that the GFC does not drive the results. While all these robustness checks underpin the insights of this paper, we cannot escape the underlying (fortunate) problem that crises are rare. Results, therefore, have to be interpreted with some caution.

These formal statistical tests confirm the insights from the raw data and previous work.

The total DSR and the credit-to-GDP gap, two indicators traditionally used as BIS EWIs, have the highest AUCs across all forecast horizons (Table 1, top two rows). While there is no statistically significant difference between their information content, the aggregate DSR has the highest AUC for the short horizon and the credit-to-GDP gap the highest AUC for the longer one. This confirms earlier findings [Drehmann and Juselius (2012)]. In line with the results from Figure 1, the property price gap performs particularly well around two years before crises, but it becomes uninformative in the pre-crisis year, when it tends to decline or close.

Household debt indicators, in particular the household DSR, are also informative (Table 1, fourth and fifth rows). Based on the AUC point estimates, the household DSR performs even slightly better than the aggregate credit gap in the pre-crisis year. It also outperforms the household credit-to-GDP gap, which we will, therefore, not consider in the rest of this article.[15]

[15] Strictly speaking, the household credit-to-GDP gap performs marginally better than the household DSR for quarters 10 to 12. These differences are not statistically significant. Still, we drop the household credit-to-GDP gap because it becomes uninformative in the pre-crisis year.

Confirming what policymakers have long stressed, international debt also contains useful information (Table 1, last two rows), although on balance not as much as the aggregate and household debt indicators. AUCs for the cross-border claims indicator are statistically significant but lower than those of the top-performing indicator, even though statistically it is hard to distinguish between the two. The foreign currency debt indicator does not perform as well as the traditional indicators throughout. To simplify the analysis, in what follows we retain only the indicator based on cross-border claims.

We next operationalize the indicators for policymaking, based on the chosen threshold criteria (Table 2). We show the values of the noise-to-signal ratio for different indicators subject to predicting correctly two-thirds of the crises. The left- and right-hand panels show the EWIs' performance over the longest available sample and over a smaller common sample, respectively.[16]

The analysis confirms that the household DSR adds value. It has the lowest noise-to-signal ratio across all indicators and samples. A 1.4 percentage point positive deviation of the household DSR from its long-run average captures around 70% of crises with a noise-to-signal ratio of roughly 20% across the two samples, i.e., one false crisis call for every five correct ones. This result is not only driven by the GFC: the household sector DSR also exceeded this threshold in four out of the six crises before 2000.

In terms of noise-to-signal ratio, the performance of the cross-border claims indicator is roughly equivalent to that of the credit-to-GDP gap, regardless of the sample considered. However, the credit-to-GDP gap predicts a larger percentage of crises.

The comparison of noise-to-signal ratios should not, however, be overemphasized. For instance, the somewhat higher noise-to-signal ratio of the credit-to-GDP gap is mainly due to its tending to signal crises very early, some five to seven years ahead of the event [Drehmann et al. (2011)]. While these are "wrong" signals according to our formal criteria, they nevertheless still correctly identify the build-up of vulnerabilities.[17]

Table 2 also highlights the EWIs' robustness. Despite large differences in sample size between the longest and the smaller common sample (left-hand panel versus right-hand panel), the thresholds for each indicator are

Table 2: Optimal signaling thresholds for standalone EWIs

EWI	ALL AVAILABLE DATA[1]				COMMON SAMPLE[2]			
	Threshold[3]	Predicted[4]	NTS[5]	# crises[6]	Threshold[3]	Predicted[4]	NTS[5]	# crises[6]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	At least 66% of the crises predicted							
HOUSEHOLD DSR	1.4	68.4	18.7	19	1.4	68.4	22.7	19
TOTAL DSR	1.8	67.9	22.4	28	1.8	68.4	24.2	19
CREDIT-TO-GDP GAP	9.0	80.0	25.7	30	9.0	78.9	32.6	19
CROSS-BORDER CLAIMS TO GDP	34.0	72.4	27.3	29	34.0	68.4	31.2	19

[1] Results are based on the broadest data coverage for each indicator.
[2] Results are based on a common sample, when all four indicators are available.
[3] Optimal EWI threshold that minimizes the noise-to-signal ratio while capturing at least 66% of the crises. In percentage points.
[4] Percentage of correctly predicted crises. A crisis is judged to be correctly predicted if the indicator variable breaches the critical threshold (column 1 or 5) anytime within a three-year horizon before a crisis.
[5] Noise-to-signal (NTS) ratio; fraction of type II errors (the threshold is breached but no crisis occurs within the next three years) divided by the fraction of correctly predicted crises (column 2 or 6). [6] Number of crises included in the analysis.

Sources: IMF, International Financial Statistics; national data; BIS credit to the non-financial sector, debt securities, locational banking and property price statistics; authors' calculations.

[16] Tables A2–A6 in the online appendix, available from the authors, show the results from a broader range of thresholds in addition to the one that minimises the noise-to-signal ratio subject to predicting 66% of crisis.
[17] Regardless of the sample, Table 2 identifies a critical threshold for the credit-to-GDP gap equal to 9 for the requirement of predicting at least 66% of the crises. This is fully in line with previous findings. It is also consistent with the Basel III calibration, which suggests that the countercyclical capital buffer should be at its maximum if the credit to GDP gap exceeds 10 [BCBS (2010)].

identical in both cases. This shows that the results are not solely due to advanced economies or crises related to the GFC – two key features of the common sample. The main insights from the table are also robust to performing bivariate comparisons for each possible pair of indicators.

5. COMBINED INDICATORS

Previous work has shown that combining information from credit and asset markets into composite indicators can improve performance. This is intuitive as financial booms feature both exuberant credit growth and buoyant asset prices. Given the role of housing as collateral, the literature has highlighted in particular how residential property prices amplify the financial cycle, despite their inferior performance as standalone indicators (Table 1).

Thus, we next derive optimal thresholds for combinations of debt variables and property prices. We follow the same logic as before. But for a warning signal to be issued we now require that (i) the debt indicator has breached the threshold and (ii) the property price gap was above 11 within the 12 quarters preceding the breach. We choose 11 because it is the standalone critical threshold obtained for this variable based on predicting at least two thirds of the crises.[18]

Table 3: Optimal signaling thresholds of combined EWIs: credit and residential property prices

EWI	THRESHOLD[1] (1)	PREDICTED[2] (2)	NTS[3] (3)
HOUSEHOLD DSR	0.5	68.8	12.1
CREDIT-TO-GDP GAP	4.0	68.2	20.9
TOTAL DSR	0.4	66.7	11.8
CROSS-BORDER CLAIMS TO GDP	19.0	70.0	19.0

[1] Optimal threshold that minimizes the noise-to-signal ratio while capturing at least 66% of crises. In percentage points.
[2] Percentage of correctly predicted crises.
[3] Noise-to-signal (NTS) ratio; fraction of type II errors (the EWI signals a crisis but no crisis occurs within the next three years) divided by the fraction of correctly predicted crises (column 2).
[4] The percentage of correctly predicted crises and the noise-to-signal ratio are based on joint signals that also incorporate information from property price gaps. A crisis is judged to be correctly predicted if (i) the debt indicator variable breaches the critical threshold (column 1) anytime within a three-year horizon before a crisis and (ii) the property price gap exceeds 11 at some point in the 12 quarters preceding the breach of the debt indicator variable.

Sources: IMF, International Financial Statistics; national data; BIS credit to the non-financial sector, locational banking and property price statistics; authors' calculations.

The condition for property prices is deliberately backward-looking. As discussed above, property price growth tends to slow from very high rates ahead of crises, so that the gap closes (Figure 1, right-hand panel).[19] If we were to require that both credit and property price gaps exceed critical thresholds simultaneously, the combined signal would start to "switch off" in the late stages of the boom.

Combining information from credit and property markets improves the EWIs' precision considerably (Table 3). Noise-to-signal ratios fall below 21%, to as low as 11.8%.

The combined EWIs also lead to lower critical thresholds for the debt indicators. This is intuitive, since the information contained in property prices underscores the signal issued by rapid credit expansion, so that the threshold can be lower.

6. ASSESSING CURRENT VULNERABILITIES

What do the EWIs suggest about current vulnerabilities? Table 4 takes a closer look at the status of the various indicators as of June 2017, while Section 6.1 provides a short discussion of how to use and interpret EWIs more generally.

The color coding is based on the standalone indicators (Table 2). Cells are marked in green if the indicator has breached the threshold for predicting at least two-thirds of the crises. Those marked in amber correspond to the lower threshold required to predict at least 90% of the crises.[20] This avoids a false sense of precision and captures the very gradual build-up in vulnerabilities. Asterisks indicate that the corresponding combined credit-cum-property price indicator (from Table 3) has breached its critical threshold.

The picture that emerges is a varied one.

Aggregate credit indicators point to vulnerabilities in several jurisdictions (Table 4, first two columns). Canada, China, and Hong Kong SAR stand out, with both the credit-to-GDP gap and the DSR flashing green. For

[18] Simultaneously changing thresholds for the debt indicators and property price gaps leads to even lower noise-to-signal ratios. But it complicates the interpretation across debt indicators for vulnerability assessments such as Table 4. As an alternative method to ensure a common property price gap across the debt indicators, we also searched for the optimal threshold for the property price gap if we minimize the average noise-to-signal ratios of the combined indicators, conditional on a common property price gap threshold for all of them. This does not deliver significantly different results.
[19] We also tried to capture the intuition from the graph by requiring not only that the property price gap is above the critical threshold in any of the previous three years but also that its current change is negative. This did not modify the forecast performance much.
[20] Provided in Tables A2–A6 in the Online Appendix, available from the authors.

Table 4: Early warning indicators for stress in domestic banking systems

	CREDIT-TO-GDP GAP	DEBT SERVICE RATIO (DSR)	HOUSEHOLD DSR	CROSS-BORDER CLAIMS TO GDP
Australia	–8.1	**1.0**	**1.0**	**21.3**
Brazil	–6.4	–0.9	…	0.5
Canada	**9.6***	**2.9***	**0.7***	**33.1***
Central and eastern Europe[1]	–12.5	–1.9	…	11.4*
China	**16.7**	**5.1**	…	–27.9
Finland	–5.2	**0.9**	**0.8**	–22.4
France	**4.0**	**1.6**	0.5	2.7
Germany	–2.1	–1.6	–1.9	6.9
Hong Kong SAR	**30.7***	**6.9***	…	–12.3
India	–7.8	**0.5**	…	–30.6
Indonesia	**6.9***	**0.5***	…	–10.9
Italy	–18.0	–1.2	0.0	–10.8
Japan	**7.6***	–1.8	–1.0	**20.5***
Korea	–1.3	**0.1**	**1.7**	–13.9
Malaysia	**4.0***	**0.4***	…	–1.6
Mexico	**6.2**	**1.1**	…	17.9
Norway	–1.0	–0.3	**1.1**	**34.4**
Russia	–4.6	**1.8**	**1.8**	–24.9
South Africa	–2.2	0.0	–0.3	**22.1**
Spain	–50.7	–3.6	–1.6	–19.5
Sweden	–11.2	–0.4	**1.1***	–1.9
Switzerland	**16.3***	**1.7***	…	7.5
Thailand	**6.7***	–0.6	**2.8***	–16.6
Turkey	**5.4**	**6.1**	…	–1.2
U.K.	–17.7	–1.3	–0.8	0.6
U.S.	–6.9	–1.1	–1.5	–15.2
LEGEND	Credit/GDP gap≥9	DSR≥1.8	DSR≥1.4	XB claims≥34
	4≤Credit/GDP gap<9	**0.1≤DSR<1.8**	**0.6≤DSR<1.4**	**18≤XB claims<34**

The threshold for green (amber) cells minimizes false alarms conditional on capturing at least two-thirds (90%) of historical crises with a cumulative three-year prediction horizon (see Table 2 and Tables A2–A6 in the Online Appendix, available from the authors). Asterisks highlight a signal of the combined indicator when property price gaps were above 11 at some point during the last three years (see Table 3).

[1] Simple average of CZ, HU and PL.

Sources: IMF, International Financial Statistics; national data; BIS credit to the non-financial sector, locational banking and property price statistics; authors' calculations.

Canada and Hong Kong, these signals are reinforced by property price developments. The credit-to-GDP gap also flashes green in Switzerland, whereas the total DSR flashes green in Russia and Turkey. Credit conditions are also quite buoyant elsewhere. Credit-to-GDP gaps and/or the total DSR send amber signals in some advanced economies, such as France, Japan, and Switzerland, as well as in several emerging market economies (EMEs). In Indonesia, Malaysia, and Thailand, as well as some other countries, property price gaps underscore this signal.

Some jurisdictions also exhibit some signs of high household sector vulnerabilities. In Korea, Russia, and Thailand, the household sector DSR flashes green (Table 4, third column). In Thailand, the green signal for the household DSR is underlined by the property price indicator. Property prices have also been in elevated in Sweden and Canada, which exhibit an amber signal for the household DSR.

The cross-border claims indicator supports the risk assessment for several countries and flags some potential external vulnerabilities for others (Table 4, fourth column). The indicator flashes green for Norway and is amber for a number of economies.

While providing a general sense of where policymakers may wish to be especially vigilant, these indicators need to be interpreted with considerable caution (see also section 6.1). As always, they have been calibrated based on past experience, and cannot take account of broader institutional and economic changes that have taken place since previous crises. For example, the much more active use of macroprudential measures should have strengthened the resilience of the financial system to a financial bust, even if it may not have prevented the build-up of the usual signs of vulnerabilities. Similarly, the large increase in foreign currency reserves in several EMEs should help buffer strains. The indicators should be seen not as a definitive warning but only as a first step in a broader analysis – a tool to help guide a more drilled down and granular assessment of financial vulnerabilities. And they may also point to broader macroeconomic vulnerabilities, providing a sense of the potential slowdown in output from financial cycle developments should the outlook deteriorate.

6.1 What do EWIs tell us?

This section explains how to read the table that assesses current vulnerabilities based on the set of early warning indicators (EWIs). Then it explains the limitations of those indicators in the context of a broader analysis of vulnerabilities,

To interpret the table entries, it helps to understand the methodology used to derive the critical thresholds that – if crossed – lead to a warning signal. For any indicator, we start off with a large sample spanning countries and time that ideally contains as many crises and non-crisis periods as possible. After checking whether the indicator has more EWI power than a coin toss,[1] we search over a range of potential thresholds that, when breached, issue a warning signal. We judge a crisis as correctly predicted if there is a warning signal at least once in the 12 quarters preceding the crisis, i.e., if the crisis occurs anytime within the three years following the breach. If a signal is issued but no crisis occurs within that time frame, we count this as a false alarm.

We choose two different thresholds to identify amber and green "alert zones". In both cases, the calibration, drawing on historical experience, minimizes the ratio of false alarms to correct warning signals (the "noise-to-signal ratio"). But one threshold is chosen so as to predict at least two-thirds of the crises (green), and the other at least 90% (amber). The green threshold is more stringent (higher) in the sense that it is exceeded less often.

The cells also include asterisks (*). These refer to instances in which the combined behavior of the corresponding debt and property price indicators signal vulnerabilities. For this debt-cum-property price combined indicator we follow a similar logic to the one above. We keep the property price gap threshold constant at its optimal standalone value and then optimize over the debt indicator threshold, so as to capture at least 66% of crises while minimizing the noise-to-signal ratio. In other words, for a warning signal to be issued, we require that (i) the debt indicator breached the critical threshold and (ii) the property price gap was above 11 (the green threshold for the property price gap on its own) within the three years before the breach.[2] When this happens, we add an asterisk to the relevant EWI.

To interpret these signals correctly from a statistical viewpoint, a few points are worth recalling:

- Over the calibration period, there were naturally many instances in which the indicators breached the thresholds (corresponding to signals denoted by the amber, green, and * identifiers) but crises did not materialize within the following three years. The more often this happens, the higher the noise-to-signal ratio.
- This may happen because crises do not materialize at all: the indicator subsequently switches off and imbalances correct themselves. Alternatively, it may happen because the signals may occur "too early" (e.g., five or six years before a crisis), with the indicator correctly continuing to signal risks until the crisis breaks out.[3] In general, even when the indicators identify the risk of crises correctly, it is unrealistic to expect them to identify the timing with any precision.
- Noisy signals also mean that the statement "66% of crises were preceded by a breach of the EWI threshold" is not equivalent to "the crisis probability is 66% once the threshold is breached." Or putting it differently, the former statement says that "given that a crisis has occurred, the threshold was breached in 66% of the cases"; the latter means "given that the threshold is breached, a crisis occurs in 66% of the cases." The reason the two statements are not equivalent is that some breaches do not herald crises, i.e., the noise-to-signal ratio is higher than zero. In fact, in our sample and as a rule of thumb, the likelihood of a crisis emerging once the threshold for an indicator is breached is around 50%.[4]

More generally, certain caveats need to be borne in mind:

- EWIs have only two settings: "on" or "off". They do not reflect the gradual intensification of a financial boom. (The use of two thresholds is designed to capture this to some extent.)
- The exact thresholds should not be overemphasized. We have run a battery of checks and drawn on other research to make sure our economic insights are as robust as possible. But the exact optimal thresholds identified can vary by a few percentage points across specifications. Given these uncertainties, whether an indicator is just above or below a threshold is not a first-order issue for monitoring purposes.
- EWIs are based on historical relationships. Thus, structural breaks may reduce their predictive power, e.g., as a result of increased use of macroprudential

measures or changes in prudential regulation more broadly. This is only partly mitigated by evidence indicating that similar variables have displayed consistent predictive power going back to at least the 1870s [Schularick and Taylor (2012)].

- EWI thresholds are common across countries. Thus, they cannot take into account country-specific features. This is inevitable: as crises are rare events, it is not possible to calibrate the indicators with any statistical confidence based on the experience of any individual country.
- The EWIs displayed in the table are specifically designed to capture only vulnerabilities linked to the financial cycle. Other vulnerabilities that could lead to banking crises are not considered (e.g. sovereign crises owing to unsustainable fiscal positions).

Taken together, these caveats suggest that EWIs cannot be analysed in isolation. They are best seen as a useful starting point for a more granular assessment of vulnerabilities.[21]

7. CONCLUSION

This article has formally assessed the performance of household and international debt as EWIs for banking distress. These variables are found to contain useful information about banking system vulnerabilities, similar to that of their more widely used counterparts based on aggregate debt. Within the group of household-based indicators, the household debt service ratio stands out. Within that of international debt indicators, cross-border claims perform better than foreign currency debt.

At the same time, in assessing these results it is also important to take into account data limitations. Crises are rare events even in samples where data coverage is good. And they become "rarer" for samples over which we can consider household or foreign currency debt. This prevents a more detailed, robust analysis of EMEs in particular. More definite comparisons and inferences would require overcoming these limitations. Thus, improving the data is an area that deserves greater attention.

[21] (1) Formally, we test whether the AUC is statistically significantly different from 0.5.
(2) We use backward-looking information for residential property prices, as the associated gaps tend to close ahead of crises (Figure 1).
(3) For instance, this is the case for the credit-to-GDP gap [Drehmann et al. (2011)].
(4) The derivation of how likely a crisis is given an EWI signal is much more sample-dependent than the thresholds shown in Table 2 because of small sample issues.

REFERENCES

Avdjiev, S., R. McCauley, and P. McGuire, 2012, "Rapid credit growth and international credit: challenges for Asia," in Pontines, V., and R. Siregar (eds), Exchange rate appreciation, capital flows and excess liquidity: adjustment and effectiveness of policy responses, The SEACEN Centre, Chapter VI. Also published as BIS Working Papers, no 377, April

Bank for International Settlements, 2017, "87th Annual Report," June

Bank of Canada, 2017, "Financial system review," November

Bank of England, 2017, "Financial Stability Report," issue no 42, November

Basel Committee on Banking Supervision, 2010, "An assessment of the long-term economic impact of stronger capital and liquidity requirements," August

Borio, C., 2014, "The financial cycle and macroeconomics: what have we learnt?," Journal of Banking & Finance 45, 182–198. Also available as BIS Working Papers, no 395, December 2012

Borio, C., and M. Drehmann, 2009, "Assessing the risk of banking crises – revisited," BIS Quarterly Review, March, 29–46

Borio, C., and P. Lowe, 2002a, "Asset prices, financial and monetary stability: exploring the nexus," BIS Working Papers, no 114, July

Borio, C., and P. Lowe, 2002b, "Assessing the risk of banking crises," BIS Quarterly Review, December, 43-54

Borio, C., R. McCauley, and P. McGuire, 2011, "Global credit and domestic credit booms," BIS Quarterly Review, September, 43–57

Bruno, V. and H. S. Shin, 2015, "Cross-border banking and global liquidity," Review of Economic Studies 82:2, 535-564

Caruana, J., 2010, "The challenge of taking macroprudential decisions: who will press which button(s)?," speech at the 13th Annual International Banking Conference, Federal Reserve Bank of Chicago in cooperation with the International Monetary Fund, Chicago.

Chui, M., I. Fender, and V. Sushko, 2014, "Risks related to EME corporate balance sheets: the role of leverage and currency mismatch," BIS Quarterly Review, September, 34–47

Committee on the Global Financial System, 2012, "Operationalizing the selection and application of macroprudential instruments," CGFS Papers, no 48, December

Detken, C., O. Weeken, L. Alessi, D. Bonfim, M. Boucinha, C. Castro, S. Frontczak, G. Giordana, J. Giese, N. Jahn, J. Kakes, B. Klaus, J. Lang, N. Puzanova, and P. Welz, 2014, "Operationalizing the countercyclical capital buffer: indicator selection, threshold identification and calibration options," ESRB Occasional Papers, no 5.

Drehmann, M., C. Borio, and K. Tsatsaronis, 2011, "Anchoring countercyclical capital buffers: The role of credit aggregates," International Journal of Central Banking 7:4, 189–240

Drehmann, M., C. Borio, and K. Tsatsaronis, 2012, "Characterizing the financial cycle: don't lose sight of the medium term!," BIS Working Papers, no 380, June

Drehmann, M., C. Borio, L. Gambacorta, G. Jimenez, and C. Trucharte, 2010, "Countercyclical capital buffers: exploring options," BIS Working Papers, no 317, July

Drehmann, M., A. Illes, M. Juselius, and M. Santos, 2015, "How much income is used for debt payments? A new database for debt service ratios," BIS Quarterly Review, September, 89–103

Drehmann, M., and M. Juselius, 2012, "Do debt service costs affect macroeconomic and financial stability?," BIS Quarterly Review, September, 21–35

Drehmann, M., and M. Juselius, 2014, "Evaluating early warning indicators of banking crises: satisfying policy requirements," International Journal of Forecasting 30:3, 759–780. Also published as BIS Working Papers, no 421, August 2013

Drehmann, M., M. Juselius, and A. Korinek, 2017, "Accounting for debt service: the painful legacy of credit booms," BIS Working Papers, no 645, June

Drehmann, M., and K. Tsatsaronis, 2014, "The credit-to-GDP gap and countercyclical capital buffers: questions and answers," BIS Quarterly Review, March, 55–73

European Central Bank, 2017, Financial stability review, May

Gourinchas, P-O., and M. Obstfeld, 2012, "Stories of the twentieth century for the twenty-first," American Economic Journal, Macroeconomics 4:1, 226–265

International Monetary Fund, 2017, Global Financial Stability Report, October, Chapter 2

Jordà, O., M. Schularick, and A. Taylor, 2016, "The great mortgaging: housing finance, crises and business cycles," Economic Policy 31:85, 107–152

Kindleberger, C., 2000, Manias, panics and crashes, fourth edition, Cambridge University Press

Laeven, L., and F. Valencia, 2012, "Systemic banking crises database: an update," IMF Working Papers, no 12/163, June

Lawrence, M., P. Goodwin, M. O'Connor, and D. Önkal, 2006, "Judgmental forecasting: a review of progress over the last 25 years," International Journal of Forecasting, no 22, 493–518

Lo Duca, M., and T. Peltonen, 2013, "Assessing systemic risks and predicting systemic events," Journal of Banking & Finance 37, 2183–2195

Lo Duca, M., A. Koban, M. Basten, E. Bengtsson, B. Klaus, P. Kusmierczyk, J. Lang, C. Detken, and T. Peltonen, 2017, "A new database for financial crises in European countries," European Central Bank Occasional Papers, no 194, July

Lombardi, M., M. Mohanty, and I. Shim, 2017, "The real effects of household debt in the short and long run," BIS Working Papers, no 607, January

Mian, A., A. Sufi, and E. Verner, 2017, "Household debt and business cycles worldwide," Quarterly Journal of Economics 132:4, 1755–1817

Minsky, H., 1982, Can "it" happen again?: essays on instability and finance, M. E. Sharpe

Önkal, D, M. Thomson, and A. Pollock, 2002, "Judgmental forecasting," in Clements, M., and D. Hendry (eds), A companion to economic forecasting, Blackwell

Schularick, M., and A. Taylor, 2012, "Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870–2008," American Economic Review 102:2, 1029–1061

Terrones, M., A. Kose, and S. Claessens, 2011, "Financial cycles: What? How? When?," IMF Working Papers, no 11/76, April

Zabai, A., 2017, "Household debt: recent developments and challenges," BIS Quarterly Review, December, 39–54

SUPRANATIONAL SUPERVISION OF MULTINATIONAL BANKS: A MOVING TARGET

GIACOMO CALZOLARI | European University Institute, University of Bologna, and CEPR
JEAN-EDOUARD COLLIARD | HEC Paris
GYÖNGYI LÓRÁNTH | University of Vienna and CEPR

ABSTRACT

Multinational banks (MNB), while potentially contributing to the efficiency of the banking sector, represent a challenge from a regulatory and supervision perspective, predominantly due to coordination failures among national supervisors. The institutional response, spurred by the recent global financial crisis, has been broad and profound, with very ambitious and substantial developments, such as the Single Supervisory Mechanism of the European Banking Union. These developments represent major changes in the organization of cross-border banking supervision and will have profound implications for the industry, worldwide. In this paper, we investigate whether and how the banking system itself will strategically adapt to such a drastic overhaul of the supervisory architecture. We illustrate cases in which this endogenous reaction of the industry may lead to unintended and probably unexpected consequences, such as higher costs for deposit insurance funds and negative impact on welfare. Although addressing coordination failures is necessary in cross-border banking supervision, we claim that policy reforms should anticipate and track MNBs' reactions, a moving target.

Multinational banks (MNB) have grown in number, importance, and complexity over the past two decades and the 2007-08 financial crisis has only temporarily slowed this process.[1] The associated rapid increase in cross-border banking claims had been seen as one dimension of the desirable integration of international banking markets. However, while banks became more international in life, they remained national in death, as the crisis has shown, with consequences that could not be contained within national borders. The financial crisis has vividly shown that MNBs active in several countries and with complex networks of foreign affiliates pose particular challenges for bank supervisors that are confined to national borders. For example, national authorities of Belgium, France, Luxembourg, and the Netherlands supervised the Belgian bank Dexia whose catastrophic failure resulted in a €6 bn bail-out in 2011. In addition, differences in institutional arrangements across countries showed up dramatically during the crisis. For example, the Icelandic bank Landsbanki operated in the U.K. with branches, whilst the competitor Kaupthing operated with a subsidiary. When the Landsbanki failed during the crisis, U.K. depositors lost some of their savings because the Icelandic Deposit Guarantee Scheme quickly became unable to reimburse all insured depositors in the country and in foreign branches of Icelandic banks. Kaupthing depositors, on the other hand, were insured by the U.K.'s

[1] See, for example, The European Economy, "Single supervisor and cross border banking," 2015 Issue n.3, https://bit.ly/2NDRicB

Deposit Insurance Scheme and were fully repaid.

The financial crisis of 2007-08 made clear that there was an urgent need to design a more comprehensive and coordinated regulatory framework in general, and for the supervision of multinational banks in particular. The institutional response has been broad and profound. For example, the Financial Stability Board (FSB) issued in 2011 an in-depth document on "Global adherence to regulatory and supervisory standards on international cooperation and information exchange." In December 2012, the Federal Deposit Insurance Corporation (FDIC) and the Bank of England issued a joint paper on international coordination of MNBs regulation and supervision. But probably the most substantial development has been the launch of the Single Supervisory Mechanism (SSM) of the European Banking Union (EBU). The SSM is an ambitious task in institutional harmonization and supervisory centralization. It contemplates several dimensions: defining and implementing a coherent harmonized legal and regulatory framework for banks; building an effective central supervisory apparatus and defining its legal framework, governance, and procedures; coordinating the operations of national competent authorities within a single rule-book in a coherent arbitrage between the union and national legal frameworks; a consolidated assessment of the balance sheet of the 129 largest and systemically relevant banks in Europe (i.e., the "most significant entities," which account for around €25 tln in assets, 80% of the Euro Area); and addressed coordination failures by giving supervision powers on these banks to a supranational authority, the European Central Bank (ECB).

These developments represent major changes in the organization of cross-border banking supervision that will have profound implications for the industry worldwide. What should not be overlooked is the fact that the banking system itself will not remain indifferent to such a drastic overhaul of the supervisory architecture.

In a forthcoming article in the Review of Financial Studies, entitled "Multinational banks and supranational supervision," we investigate the effects of centralization of supervision of MNBs within a supranational authority, taking account of the industry's adjustments.

Putting a supranational supervisor, such as the ECB, in charge of all of an MNB's units solves the issue of "supervisory failing" that became evident during the crisis. A single supervisor, internalizing the consequences of its actions across countries, is the type of international coordination that was lacking during the crisis. However, as stated, the industry itself will not be indifferent and will very probably adjust to these major institutional changes in ways that will make the overall effects of supranational supervision far less obvious.

To address these possibilities, we investigate supervisors' incentives to monitor the activities of a bank. The supervisor can detect poorly performing assets and steer the bank towards safer activities, thereby reducing the risk of failure and the costs to the deposit insurer. The incentives to monitor clearly depend on the different institutional environments and the bank's network organization of foreign affiliates. Banks can expand abroad with independently incorporated subsidiaries that are protected by limited liability with respect to the other banks within the cross-border group. The current institutional arrangements imply that, absent international coordination, the host authority is in charge of supervision of foreign subsidiaries and in case of failure the host authority will be responsible for reimbursing local depositors. Alternatively, MNBs may expand abroad with branches that share assets and liabilities with the "mother" bank in the "home" country of origin (because they are not independent legal entities) and are normally supervised and insured by the home country authorities. If supranational supervision comes into effect, such as with the SSM, the responsibility of supervision goes to the supranational authority, such as the ECB, irrespective of the bank's organization vis-à-vis its subsidiaries and branches. Deposit insurance could in principle be centralized and mutualized as well. However, to our knowledge this step has not been followed yet, neither in the SSM nor in other cases of coordinated supervision.

In our analysis, we clearly show how the lack of coordination generates undesirable consequences in terms of externalities, i.e., material effects of action or inaction of a national authority on the other countries in which the MNB is active. In particular, if the foreign supervisor decides to limit the monitoring of the foreign subsidiary of the MNB, it reduces the availability of information about the foreign subsidiary's assets that could be important for the supervisor of the home unit.[2]

[2] Foreign assets could be seized, at least in part, by the home supervisor in case of trouble with the home unit of the MNB. The other direction of the externality (e.g., the home supervisor limiting monitoring) is less of an issue because the liability structure of a subsidiary is asymmetric, as explained, and the foreign supervisor knows that it cannot expect much from the home units in case the foreign subsidiary faces difficulties.

Furthermore, monitoring, and subsequent possible conservatorship interventions in the foreign subsidiary, directly reduce the availability of these (or some of these) foreign assets to counter losses in the home unit.[3] None of these effects would be present in case of an MNB organized with foreign branches (the home supervisor would be responsible for monitoring and insuring all depositors worldwide) or in the case of supranational supervision independent of the organizational form of the foreign unit.

Investigating these two externalities, we are able to show that supranational supervision does not necessarily imply more intense monitoring than in the case of independent national supervision, as the two externalities interact in a subtle way (more details can be found in our paper). However, if the bank does not modify the organizational form of its foreign activities, we can predict that supranational supervision reduces the expected costs for deposit insurance funds, which is a desirable effect of coordinated supervision.

What is more, we are able to go one step deeper into our understanding of supranational supervision and predict whether and how the MNB may want to adjust its foreign organization to the modified supervisory and monitoring environment.

Since the bank's expected profits are depressed by monitoring and associated risk-reducing interventions, when supranational supervision implies more monitoring of foreign subsidiaries the bank has an incentive to reorganize into foreign branches. Alternatively, if this reorganization involves excessive administrative costs, it may even prefer to shut down foreign activities altogether, thus reducing the internationalization of banking markets. The opposite holds in case of reduced monitoring with supranational supervision.

Interestingly, in either case the MNB readjusts towards foreign organizations and activities in pursuit of reduced monitoring. Both the reorganization of banking activities and the associated reallocation of monitoring then have material consequences on the (expected) costs of deposit insurance funds and overall welfare. This is different from what is found if the MNB keeps its foreign organization unchanged. When it does adjust its structure to supranational supervision and transforms foreign branches into subsidiaries (or shutting down foreign branches), the expected costs of deposit insurance funds increase and overall welfare reduces. Surprisingly, national supervision in this case is a superior institutional arrangement than supranational supervision. More benign consequences are realized instead when the bank transforms the foreign subsidiaries into branches or opens foreign units, as a consequence of supranational supervision. Moreover, all these changes in the bank's activities have material consequences not only on the actual costs for the deposit insurance funds but also on their distributions across countries.

Are MNBs effectively adjusting to the moves towards coordination and supranational supervision, such as the SSM in Europe? It is probably too early to say for sure, but there are important signs that this is already happening. For example, the large Scandinavian bank Nordea in 2017 converted its subsidiaries in Denmark, Finland, and Norway into branches and also announced the move its headquarters from Sweden to Finland, thus ultimately transferring supervisory responsibility to the SSM. At a more general level, in the Euro area countries there were 550 branches and 310 subsidiaries from other E.U. countries in 2012.[4] Although the number of branches remained almost unchanged in 2016 (552), that of subsidiaries dropped to 232. At the global level, the number of subsidiaries have instead increased over the same period,[5] which leads one to think that the marked different trend and the relative "branchification" in the Euro area is (also) a consequence of SSM and the inception of supranational supervision.

Note that, luckily, the changes observed in the case of the SSM in Europe correspond to benign effects of supranational supervision and industry reaction (from subsidiaries to branches). However, it is important to point out that despite the intense policy debate that took place when the SSM was planned, there was no explicit discussion of how cross-border banks could have responded to this institutional change, at least to our knowledge.

In the end, it seems that the industry is reacting in a positive manner, but things could have gone (and can still go) wrong with transformation of foreign branches

[3] The home supervisor is, in any case, a residual claimant of foreign assets, if the foreign subsidiary has enough residual assets and after foreign depositors have been reimbursed. Consequently, it cares for the upside of the returns in the foreign subsidiary.

[4] See the ECB's report on the E.U. structural financial indicators, https://bit.ly/2xBYHij

[5] See, The globalisation of banking: How is regulation affecting global banks?", BBVA Research, 8 August 2016.

into subsidiaries (as it happened, for example, with the branches of the Greek Alpha Bank in Romania and Bulgaria that were bought in 2015 by other Greek banks and then reorganized into subsidiaries, backed up by the more solid Romanian and Bulgarian national deposit).[6]

We think our analysis provides a broader message that is reminiscent of the famous Lucas' critique in macroeconomic policies of the seventies. If we want to understand and predict the consequences of important policy shifts, such as the move to supranational supervision, we need to understand how the relevant economic agents, in our case banks, will react and should not take the environment, the banking industry structure, as given and fixed. The strategic reaction of agent, the banks, can be substantial and, in some cases, unwind the intended consequences of policy reforms. This need for "microfunding" policy reforms by accounting for banks' reaction is in fact a general message that goes beyond the case of supranational supervision. For example, the SSM currently lacks a European common deposit insurance that is a major limitation of the Banking Union architecture. However, we need to understand the potential reaction of the European banking industry to such a reform and its redistributive consequences across countries.

We conclude by mentioning a new policy perspective that emerged in our research as a sensible and effective one when dealing with cross-border banking, supervision, and associated externalities. An MNB organized with foreign subsidiaries effectively enjoys an implicit subsidy as compared with foreign branches. In fact, the foreign deposit funds may have to reimburse foreign depositors even if other units in the group are paying dividends to shareholders. To account for this implicit subsidy, one could account for the different costs that different bank organizations (branch or subsidiaries) impose on deposit insurance funds, to some extent in line with the idea of risk-based insurance premia. This type of "representation-form-dependent" premia, would align the bank's profitability with welfare and provide banks with the incentives to adopt a preferred representation form that reflects welfare at large. With this type of policy, the moving target of supranational supervision, i.e., the reorganization of banks, would not be a problem because the policy itself adapts to the target.

[6] See "Greek Eurobank Takes Over Alpha Bank's Branch Network in Bulgaria," July 18, 2015, at www.novinite.com.

REFERENCES

Calzolari, G., J.-E. Colliard, and G. Loranth, 2018, "Multinational banks and supranational supervision," forthcoming, Review of Financial Studies, available at https://bit.ly/2PZbMZK

FINANCIAL STABILITY AS A PRE-CONDITION FOR A HARD BUDGET CONSTRAINT: PRINCIPLES FOR A EUROPEAN MONETARY FUND

DANIEL GROS | Director, CEPS[1]

ABSTRACT

Since the financial crisis mutated into a 'euro' crisis in 2009-10, the feasibility and desirability of creating a European Monetary Fund (EMF) has been the object of serious debate in both academic and policy circles. In the meantime, the European Stability Mechanism (ESM) was created to essentially perform the functions of an EMF. It has been critical in containing the cost of the crisis, and four of its five country programs have been a success. But the case of Greece shows that one needs to be prepared for failure as well. This contribution proposes to keep the ESM's remit essentially as it is today but further empower it to impose conditions on countries receiving its financial support. Such support, however, would be limited to prevent situations arising in which the ESM would come to 'own' a country.

Within such a structure, the ESM/EMF is viewed literally as a financial stability mechanism, whose main function is to ensure that a bailout is no longer 'alternativlos', as Chancellor Angela Merkel used to say. In 2010, the rescue of Greece was presented as TINA (There Is No Alternative) because the stability of the financial system of the entire euro area appeared to be in danger. With financial stability guaranteed by the ESM/EMF in combination with the Banking Union, default becomes an alternative that should be considered dispassionately. Whether the debt of a country is sustainable can rarely be known with any certainty beforehand. Accordingly, it is proper that the Union, in the 'spirit of solidarity', initially gives a country the benefit of a doubt and provides financial support for an adjustment program. But the exposure of the Union should be limited. If the program does not work as expected, a hard budget constraint needs to be imposed, but the ESM/EMF could still be of great help as a source of bridge financing to soften the cost of default.

1. INTRODUCTION

The idea of a European Monetary Fund (EMF) has once again become fashionable. European Commission President Jean-Claude Juncker became a convert in his 2017 State of the Union speech, and the full Commission endorsed this concept later in its 'St Nicolas' package of proposals to improve the governance of the euro area. Several German policymakers subsequently also called for a transformation of the European Stability Mechanism (ESM) into an EMF. Franco-German negotiations started in this direction in 2018.

However, the various advocates of an EMF have very different ideas about its purpose and functions. This paper looks at the essential functions of an EMF and asks what changes would be needed to the ESM in order to improve the functioning of the euro area.

It will be useful to start by offering a short background of this idea. This article then turns to an examination of

[1] This contribution is based on the author's earlier publication, co-authored with Thomas Mayer, entitled "A European Monetary Fund: why and how?" CEPS Working Document No. 2017/11, CEPS, Brussels, December 2017.

the ESM's performance so far and sketches the (limited) changes that might be necessary to allow it to take on this expanded role.

2. THE HISTORY

When the first proposal for an EMF was published [Gros and Mayer (2010b)], in February 2010, Greece was still struggling on its own to avoid default. Following the revelation of a much higher government budget deficit in 2009 than had been expected earlier, 10-year government bond yields had increased from 4.5% in August 2009 to 6.1% in January 2010. Although the prospects of Greece being able to roll forward maturing debt in the market were slim, the proposal met with widespread rejection. Most people felt that the E.U. institutions would be unable to agree on financial support for a country at risk of default in view of the no-bailout clause enshrined in the Maastricht Treaty. Moreover, treaty change was dismissed as impossible, and the Gros and Mayer (2010b) proposal was seen at best as a project for the distant future.

Two months later, however, things were moving very quickly. On Sunday, May 2, 2010, Greece received its first support program, with the funds coming from bilateral loans from other Economic and Monetary Union (EMU) countries. But the move failed to calm markets, and market participants lost confidence in the liquidity and solvency of other EMU countries. This prompted the European Council (in this case the meeting of the heads of state and government of the euro area members) on the following Sunday, May 9, to create a €500 billion fund dubbed the European Financial Stability Facility (EFSF) to be able to give support to a broader group of countries.

Because the EFSF needed time to be organized, on the same day the European Central Bank (ECB) launched a government bond purchasing program, called the Securities Markets Program, with a view to bolstering the sagging prices of bonds issued by euro area governments. The EFSF was originally intended to be temporary, but the evolution of the crisis, with Portugal and Ireland needing funding (and the Greek program failing), showed that there was a need for a permanent structure to help countries in temporary financial difficulties. The EFSF was then de facto merged into the permanent ESM in late 2012. In the summer of that year, the crisis seemed to spread to two large countries, Italy and Spain. At this point, it appeared that the entire euro area was in danger of dissolving. This prompted Mario Draghi, President of the ECB, to assert that his institution would do "whatever it takes" to prevent a disintegration of the euro. Financial market tensions rapidly abated and the ECB replaced the Securities Markets Program with the Outright Monetary Transaction (OMT) Program in September 2012.

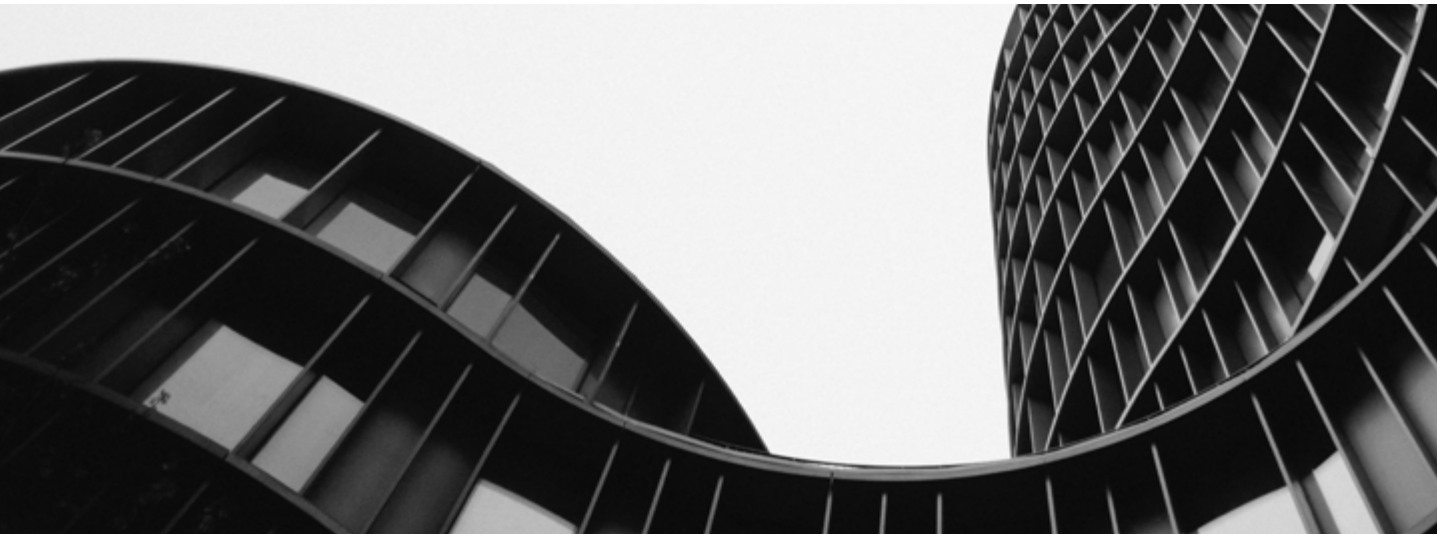
The OMT has since been widely credited with providing an indispensable safety net. But in reality, the activation of the OMT (under which the ECB would buy only short-term government bonds) is subject to the conclusion of an ESM program. The purpose of the OMT was not to substitute the ECB for the ESM, but rather to ensure the credibility of the euro area as a whole when the stability of the entire area is in danger. In such a situation, the resources of the ESM would clearly be insufficient. The International Monetary Fund (IMF) is likewise responsible for providing adjustment financing for individual countries, but its resources would not be sufficient to deal with a European financial crisis. In short, while in early 2010 an EMF seemed to be utopic, it has de facto come into existence since late 2012.

3. THE ORIGINAL IDEA

The initial blueprint [Gros and Mayer (2010b)] was, of course, very much influenced by what was then regarded as the key problem, namely the deterioration of economic conditions in Greece. We thus concentrated on five main issues, outlined below.

1. Financing mechanism: it was envisaged that capital contributions to the EMF would be based on the potential risk a country represents to the EMU. Hence, we proposed that countries breaching the Maastricht criteria would make higher contributions based on the excess of their public debt and deficit ratios above 60% and 3% of GDP, respectively. We expected that our mechanism would create a capital base of €120 billion over time, which could be leveraged to a funding capacity of at least €500 billion through borrowing.

2. Conditionality: this aspect was conceived to consist of two stages. In stage I, any EMF member could call on the capital (and cumulated interest) it had subscribed; possibly in the form of an EMF guarantee of new issues of public debt, provided its fiscal adjustment program was approved by the Eurogroup. In stage II, use of assistance from the EMF greater than the capital subscription would be dependent upon a tailor-made adjustment program supervised by the European Commission and the Eurogroup.



3. Enforcement: if a country did not live up to its commitments, new financial assistance would be cut off. A continuing breach of conditions would lead to a cut-off from structural funds and, in the event, from the euro area's money market, as the public debt of the offending country would no longer qualify as collateral for ECB funds under a repurchase agreement.

4. Orderly default: it was deemed important to recognize that default was possible, but also that the costs of default would have to be contained. If a country could not make the necessary adjustment effort, it would be in everybody's interest to cut the debt burden. To make restructuring possible, Brady bonds were held up as a useful model to follow, in which bad debt is exchanged against safe debt backed by the EMF with a haircut.[2] The size of the haircut should be such that it would bring down the debt ratio of the country in question to the Maastricht limit of 60%. In return for exchanging bad debt against safe debt, it was suggested that the EMF would acquire all claims against the defaulting country. From that time onwards, any additional funds the country received could be used only for specific purposes approved by the EMF. Other E.U. transfer payments would also be disbursed by the EMF under strict scrutiny, or they could be used to pay down the debt owed by the defaulting country to the EMF. Thus, the EMF would provide a framework for sovereign bankruptcy comparable to the Chapter 11 procedure in the U.S. for bankrupt companies that qualify for restructuring. Without such a procedure for orderly bankruptcy, the Community could be taken hostage by a country unwilling to adjust, threatening to trigger a systemic crisis if financial assistance is not forthcoming.

5. Exit: since member states of the E.U. remain sovereign countries, the original proposal acknowledged that a defaulting country could regard such intrusion into its policies by the EMF as a violation of its sovereignty and hence unacceptable. But an E(M)U member country that refused to accept the decisions of the EMF would, of course, lose access to financing from the EMF and would then have to choose between introducing capital controls or leaving the euro. At the same time, its debt towards the EMF would continue to exist and would have to be serviced anyway. If a country refused to do this and declined all cooperation, its membership of the E.U. would be called into question.[3]

This article assesses the extent to which these concerns have been addressed following a brief analysis of the experience so far.

[2] The Brady plan offered two options: i) exiting for those investors willing to take a haircut and ii) remaining invested, but in this case also providing fresh money. In the Brady plan, the reduced principal amount was partially collateralized by specially issued U.S. Treasury 30-year zero-coupon bonds purchased by the debtor country using a combination of IMF, World Bank, and the country's own foreign currency reserves. Accordingly, the debt of the defaulting country after the haircut would be collateralized by EMF guarantees.

[3] In extreme cases, it could effectively be thrown out by recourse to Article 60 of the Vienna Convention on International Treaties, or Article 7 of the Treaty of Lisbon could be invoked.

4. THE EXPERIENCE SO FAR

The track record of the five adjustment programs the ESM (and its predecessors) has undertaken is mixed. Four of the five rescue programs have already ended and could be described as a qualified success. The financing difficulties of Portugal, Ireland, and Cyprus turned out, ex-post, to be temporary (although these countries continue to depend on low interest rates due to their high levels of debt), and the recapitalization of the Spanish banking system was not so expensive. These adjustment programs have, of course, been widely criticized as being too harsh – a criticism leveled against most IMF programs. There is no need to take a stance on this issue here. What matters is that the financing of the ESM avoided three further defaults (and in the case of the small program for Spain alleviated that fear considerably). The ESM hence performed the standard function of an EMF.

The political cost of even these 'successful' cases, however, has been high. The financial assistance packages are not remembered for what they prevented (insolvency and financial collapse), but for the perceived cost of the 'austerity' in terms of incomes, employment, and output. As William Shakespeare remarked centuries ago, lending is an ungrateful business, as the lender risks losing both the friendship and his money.[4]

The one case that is almost universally regarded as a failure is that of Greece. The country obtained considerable debt relief but still needed three programs of ever-increasing amounts, and there is little sign even today of a sustained recovery. The following analysis concentrates on the Greek case and asks what features of the ESM framework might be changed to prevent similar problems in future. Before this, we examine to what extent actual developments have been in line with our earlier recommendations.

5. SIMILARITIES AND DIFFERENCES

In comparing the original blueprint [Gros and Mayer (2010b)] with what has been created so far, a considerable number of differences emerge, as presented below.

- **Financing mechanism:** Under the original scheme, financial contributions were to be based on the potential risk a country represents to EMU (e.g., its debt level), whereas contributions to the ESM are based on countries' shares of the capital of the ECB (the simple average of the respective country's shares in the total population and GDP of the euro area). Applying the ECB's capital keys to the ESM has permitted smaller countries, which are potentially more exposed to the risk of sudden stops in cross-border financial flows, to also have smaller capital shares in the ESM. This is obviously inconsistent with the key principle of insurance, namely that contributions to a common pool should not only be based on the size of the risks covered, but also on the exposure of the insured to these risks.
- **Conditionality:** Comparable to the original two-stage model for access to funds and the associated conditionality, the ESM offers several stages of access, ranging from precautionary credit lines and the purchase of bonds of a member state in primary and secondary markets to adjustment loans. In addition, the ESM can also lend to member states for the purpose of recapitalizing insolvent banks, and under certain conditions to recapitalize these entities directly. All financial assistance by the ESM comes with policy conditions specified in a memorandum of understanding agreed with the European Commission, the ECB, and (where applicable) the IMF. So far, so good. However, the involvement of the ECB in the design and monitoring of financial assistance is problematic, as it blurs the distinction between monetary and fiscal policy. Moreover, the continuing involvement of the IMF in intra-EMU affairs cannot be taken for granted, given the institution's global mandate and shareholder base. Finally, yet importantly, the European Commission's new understanding of its role as a "political commission" (as it was described by Jean-Claude Juncker following his appointment as President in October 2014) is incompatible with the job of designing and monitoring conditional financial assistance (where political aspects ought to be minimized).
- **Enforcement:** Contrary to the original idea of strict enforcement of the conditions for financial assistance, the bodies entrusted with the monitoring of assistance (the so-called Troika, consisting of the ECB, the IMF, and the European Commission) have shown considerable leniency, although the program countries have a very different impression. Their repeated non-compliance has been met with base drift or a watering down of the benchmarks, though not all of it can be attributed to policy slippage in the program countries. Some of the benchmarks were impossible to reach

[4] "Neither a borrower nor a lender be; For loan oft loses both itself and friend." (Hamlet, edited by Thompson A., and N. Taylor, Arden Shakespeare, 3rd series, London: Thomson Learning, 2006).

because, at least at the start, the underlying economic assumptions (notably on exports and growth) were much too optimistic. In the case of Greece, for example, the result was that after much delay the country reached most fiscal targets and passed most reforms. But it also became apparent that while the Greek government and parliament could be pressured to pass all the laws and regulations demanded by the creditors, it proved impossible to reform the administration, whose inefficiency and obstruction prevented in many cases the actual implementation of intended structural reforms.

- **Orderly default:** As mentioned above, Greece benefited in March 2012 from a €107 billion debt reduction, equivalent to a 53.5% haircut on the principal value of about 97% of outstanding bonds held by private-sector creditors (€197 billion).[5] Yet, the process leading up to this was anything but orderly. The announcement made in late 2011 by the French and the German leaders that some form of haircut would be considered greatly unsettled financial markets.
- **Private-sector involvement:** The EFSF supported the restructuring in a way that was similar to the earlier Brady plan. In the so-called private-sector involvement (PSI) facility, Greece offered investors one- and two-year EFSF bonds. These EFSF bonds, provided to holders of bonds under Greek law, were subsequently rolled over into longer maturities. In the bond interest facility, Greece offered EFSF six-month bills to investors in order to enable the country to repay accrued interest on outstanding Greek sovereign bonds under Greek law that were included in the PSI. The bills were also subsequently rolled over into longer maturities. The operation largely followed the pattern originally envisaged in the 2010 blueprint. Still, although the debt reduction of €107 billion amounted to 56% of nominal GDP at the end of the second quarter of 2012, the actual debt-to-GDP ratio at the end of the year was only 12 percentage points lower than before the debt reduction (160% of GDP at the end of 2012 versus 172% at the end of 2011). This situation can be explained by several factors. First of all, the headline debt reduction of €107 billion is misleading, since the approximately €60 billion of debt held by Greek banks was nominally cut in half, but the EFSF then had to immediately lend the Greek government €30 billion to recapitalize the banks. Moreover, even in 2012, the Greek government was still running a sizeable deficit, which needed to be financed from external sources. On top of this, the "sweeteners" provided to some investors in the PSI operation also increased Greece's debt vis-á-vis the EFSF. Finally, nominal GDP continued to decline, thus increasing the debt-to-GDP ratio. To achieve a more substantial debt reduction without imposing an even-greater haircut on private creditors, official creditors would have had to participate in the exercise. Their refusal to do so was one of the reasons why the exercise failed and the sacrifice of the private creditors was made in vain.
- **Make failure possible?:** At the conclusion of their crisis meeting on 21 July, 2011, the European Council members stated: "As far as our general approach to private-sector involvement in the euro area is concerned, we would like to make it clear that Greece requires an exceptional and unique solution. All other euro countries solemnly reaffirm their inflexible determination to honor fully their own individual sovereign signature and all their commitments to sustainable fiscal conditions and structural reforms."[6] This statement may have been necessary to calm markets after the announcement of Greece's partial default, but it also seemed to close the window for any further and orderly public-debt restructuring in the euro area. Fortunately, European Council Conclusions have no direct legal power. Moreover, the statement is sufficiently ambiguous. It does not say that defaults should never occur again, only that "Greece requires an exceptional and unique solution." Future defaults will certainly be different and require a different "unique solution."
- **Exit:** As would be expected, the citizens of the crisis countries regarded the intrusion into their policies by the euro-area crisis management as a violation of their sovereignty and strongly resisted such action. Greece came close to exiting EMU on two occasions. In 2012 and 2015, the idea of creating a Greek parallel currency to the euro was considered by the Greek government. And during the negotiations of a third assistance program in 2015, the German Finance Minister proposed to the Council of Ministers a temporary exit of Greece from EMU. According to press reports, apart from Greece itself, the German proposal was opposed by France, Italy, Spain, and Luxembourg. It was finally dropped at the next meeting

[5] Bonds held by official creditors, notably bonds acquired by the ECB under its Securities Markets Program, were exempted from restructuring.

[6] European Council, 2011, "Statement by the Heads of Government or Euro Area and E.U. Institutions," 21 July, https://bit.ly/2QcM1Fz

of the European Council. But the idea of exiting the euro (or of introducing a parallel currency) has not died and is still actively discussed by major political figures in Italy, for example.

In summary, both the crisis management and design of crisis-management institutions have been overshadowed by deep – one could even say "philosophical" – differences of opinion about the role of discretionary policy and contractually agreed rules [Brunnermeier et al. (2016)]. This often resulted in poor compromises, which rendered a resolute solution of the crises more difficult and left many market participants confused and skeptical about the survival of the euro. It eventually took the de facto guarantee of the ECB to use its unlimited monetary firepower to defend the euro to calm markets. Nonetheless, there is a general consensus that the intervention of the ECB can only be of a temporary nature. To put EMU on a firm footing, more comprehensive institutional changes are required.

Experience has shown that a public-debt crisis can arise from two sources: i) overspending by the government itself and ii) a financial boom-bust cycle that leads to a deep recession and forces the government to bail out its banks. Greece and Ireland represent the two archetypal cases. In principle, there are now mechanisms that should make both less likely. The provisions of the Bank Recovery and Resolution Directive (BRRD), which require a high level of capital (approximately 8% of the balance sheet) that can be 'bailed in' before public-sector support is needed, should already drastically reduce the burden of future financial crises for public finances.[7] Moreover, the common funding for bank restructuring available from the 'Single Resolution Fund' (SRF) would further reduce, perhaps even eliminate in most cases, the need for national governments to provide financial support for their banks.[8]

The Fiscal Compact mandates a continuous reduction in debt-to-GDP ratios, which should significantly diminish the probability of future public-finance excesses. Theoretically, the need for EMF assistance should likewise diminish over time. In practice, however, implementation of the Fiscal Compact has remained patchy. We, therefore, concentrate on the analysis of the first type of crisis, hoping that it will at least have become less likely, thanks to whatever limited effect the existing fiscal rules have.

6. WHAT SHOULD BE DONE DIFFERENTLY NOW?

Given that the idea of an EMF has made its way back onto the agenda for the completion of EMU, it is worth asking what should be changed in the earlier proposal with the benefit of hindsight. Here are the main points:

- **Limiting financing:** To avoid turning a crisis-assistance facility into a scheme for permanent transfers and subsequent dependency, as de facto happened in Greece, financial assistance should be limited. The IMF has recently adopted access limits for its own lending, which imply that under ordinary circumstances financial assistance is limited to five times the quota the country has in the IMF. A similar limit seems appropriate for the ESM/EMF. That being said, IMF quotas are determined somewhat differently from those of the ESM. Quotas in the ESM are the same as those in the ECB, which are an equally weighted mix of GDP and population. This corresponds closely to the most important element in the quota formula of the IMF, which assigns a weight of 50% to GDP, 30% to openness, and 15% to economic variability. The latter two factors are very much related to country size: smaller countries are typically more open and are often more exposed to shocks for the simple reason that small economies are less diversified. To capture differences in risk exposure, one might modify the overall access rule of five times the ESM quota by increasing the 'multiplier' for small countries to seven (times the ESM quota) and reducing it also to three times for the very large countries.

  Table A1 in the Appendix shows the resulting access limits, together with actual ESM funding in the five programs that were undertaken. Had this key been applied to the program countries, the assistance to Greece would have been many times larger than the actual sum the ESM provided to the country. For Spain and Cyprus, the proposed limits would not have been a constraint. The assistance to Ireland and Portugal would have been above the limit if the IMF and ESM funding were combined. If one deducts the financial assistance used for bank recapitalization, the programs in Ireland and Portugal would have been considerably smaller and the proposed access limits would not have been binding. In the case of

[7] Directive 2014/59/EU of the European Parliament and of the Council of May 15, 2014 establishing a framework for the recovery and resolution of credit institutions and investment firms, OJ L 173, 12.6.2014.

[8] De Groen and Gros (2015) show that the SRF would have been sufficient to cover the necessary bank restructuring funding during the euro crisis, if the BRRD rules been strictly applied.

Ireland, €24 billion of the total assistance of €67.5 billion was used for bank recapitalization, leaving €43.5 billion for fiscal support (equivalent to 110% of the proposed access limit). Notably, part of the fiscal support had become necessary because the Irish government had already injected €46 billion into the banking sector. Had the government not stepped in to bail out bank creditors, it is doubtful whether Ireland would have needed financial assistance at all. In the case of Portugal, €12 billion of the total assistance of €78 billion was used for bank recapitalization, leaving €66 billion for fiscal support (106% of the proposed access limit). Thus, without the additional bank recapitalization, the proposed access limit would have allowed for providing assistance of roughly the size granted for fiscal support. As mentioned above, in future, national funding for bank recapitalization should no longer be needed given the bail-in rules of the BRRD and the SRF. The need for financial assistance to governments should be considerably reduced once the SRF becomes fully operational, which will be the case soon.

- **Conditionality:** with hindsight, the various stages of access to ESM assistance would also seem appropriate for the EMF. Responsibility for the design and monitoring of adjustment, however, would have to be assigned to a European institution that operates at arms' length from politics. The EMF would, therefore, have to develop its own capacity to monitor economic developments and implement adjustment programs (although the European Commission would continue to play a role in economic monitoring as enshrined in the EU Treaties). With access to financial assistance limited, the length of programs should also be limited to, say, three years (in line with standard IMF practice). To avoid the substitution of private debt by public debt during the program, debt service payments would need to be suspended for the duration of the adjustment program (as discussed in the Appendix). The governance structure of the ESM would seem to be appropriate to task ESM staff with program design and monitoring, and the board of directors with approval of staff decisions. The roles of the ECB, IMF, and Eurogroup would become redundant.
- **Enforcement:** enforcement of conditionality has been too weak. Hence, it is important to reiterate the need to cut off new financial assistance if a country did not live up to its commitments. A continuing breach of conditions should lead to cut-offs from structural funds and, in the event, from the euro area's money market, as the public debt of the offending country should no longer be eligible as collateral for ECB funds under a repurchase agreement (which implies that there would be capital and exchange controls). Emergency lending assistance to banks by national central banks would also be abolished.
- **Orderly default:** in line with the earlier proposal, a country would be expected to restructure its public debt if the size and length of the adjustment program is not enough to bring the country back to the market. The Brady bond model and the Greek version of it still seem appropriate, and we continue to think that the size of the haircut should be such that the debt ratio of the country in question declines to the Maastricht limit of 60%. In return for exchanging bad debt against safe debt, the EMF should receive all claims against the defaulting country. From that time onwards, any additional official funds the country received could be used only for specific purposes approved by the EMF. Other E.U. transfer payments would also be disbursed by the EMF under strict scrutiny, or they could be used to pay down the debt owed by the defaulting country to the EMF.
- **Exit:** if both financial assistance and debt restructuring failed to create financial stability and the respective country were cut off from all further assistance from the EMF, it should be able to reintroduce its own currency, exclusively or in parallel to the euro, without having to leave the E.U. The above regulation of structural funds and other E.U transfers would still apply.

The European Commission recently published its own proposal for an EMF. It differs from the earlier scheme we have examined in several respects: First, the Commission's proposal introduces a "reinforced qualified majority" voting procedure to speed up decisions, but there is no evidence that the existing voting procedure prevented timely assistance when needed. Second, the Commission intends to establish the EMF as a Community institution with strong Commission involvement instead of an intergovernmental institution. In the view of the present author, however, this is inconsistent with the character of the EMF as a non-political institution and the Commission's declared intention to be "political." Third, the Commission wants to develop new financial instruments within the EMF, but there is no scarcity of

Community financial instruments and a clear focus of the EMF on providing emergency financing for governments is decidedly preferable. Fourth, the Commission's proposal omits sovereign debt restructuring and EMU exit scheme, both of which are arguably necessary as error-correction mechanisms in a world governed by uncertainty. The Commission also wants to enable the EMF to act as a common backstop to the SRF, which would be consistent with the model of the ESM as a lender of last resort for official entities, including the SRF. In a large crisis, many assets are often underpriced, and hence it is likely that the SRF would be able to repay a loan from the EMF not only from future fees it will receive from banks but also from profits on bank assets it has acquired in the course of managing the crisis.

This revamped outline for an EMF would strengthen the incentives for establishing sound public and private finances in EMU member states and thereby reduce the reliance on the ECB to maintain the monetary union. It would also respect the principle of no bailout, which is still enshrined in the Treaty on the Functioning of the European Union, much more clearly than the present arrangements (which have been subject to numerous legal cases). Almost nothing in life is irreversible, not even entry into EMU, and defaults are a fact of life in a market economy. If one accepts this reality and prepares for the consequences, the ECB would no longer need to act as a quasi-fiscal agent and could concentrate on its original mission, namely to issue money with a stable purchasing power (or 'inner value') for the citizens of the monetary union.

> "The main purpose of an EMF should be to ensure the stability of the financial system of the euro as a whole. In this way, the cost of a default would be limited and the Union would not have to bear the cost of excessive debt accumulation by any single member."

7. CONCLUDING CONSIDERATIONS

The successful rescue programs of the ESM have shown the value of having a lender of last resort for solvent, but illiquid governments. The case of Greece has also shown the difficulty of distinguishing between solvency and liquidity.

In concluding, I would like to stress another consideration that emerges from the euro crisis. When the financial system of the entire area is in danger, governments feel that they have no choice but to bail out even governments that are very likely to be insolvent. Moreover, those same governments and the E.U. institutions will even pressure a national government with a troubled economy to accept a bailout in order to limit broader financial instability. This makes it difficult to impose conditions and increases the political costs for both creditors and debtors, as both feel that they are not acting in their own interests.

Thus, the key purpose of an EMF should be to ensure the stability of the financial system of the euro in order to limit the negative spill-overs from the potential financing difficulties of any individual member state [Tirole (2015) and Farhi and Tirole (2017)]. This is essential, not with the punitive intent of "establishing market discipline," but to ensure a proper alignment of responsibilities: the Union should not have to bear the cost of excessive debt accumulation of member states, which ultimately remain sovereign in their fiscal policies [Schäuble (2017)].

Member states will respond to proper incentives to reduce their debt to sustainable levels only if they know that the Union is not obliged to bail them out. Since financial crises spread via contagion, this implies that an EMF should have ample facilities to protect "innocent bystanders," i.e., those countries whose finances are sustainable but which might suddenly experience financing difficulties because investors withdraw from an entire group of countries. Another way in which financial crises spread is via the banking system. It is, therefore, critical that the unsustainable debt of a government does not put the Euro Area's banking system into difficulties, which is, of course, the purpose of the Banking Union. But given that the resources of the SRF are limited, it might be useful to clarify that the Union will stand behind the institutions of the Banking Union in the event that a large crisis emerges. In other words, in the existing credit money order, the Banking Union eventually needs a fiscal backstop (or the money order will need to be changed).

A financial stability mechanism is thus essential to ensure that a bailout is no longer "alternativlos," as Chancellor Angela Merkel used to say. The EMF should create the possibility to decide whether to grant financial support to a country that cannot roll over its debt because it has lost market access. It makes a world of difference whether, as in 2010, both sides feel condemned to accept a bailout package that neither likes, or whether there are alternatives. Whether the debt of a country is sustainable is rarely known with any certainty before a crisis strikes. Accordingly, it is proper that the Union, in the 'spirit of solidarity', initially gives a country the benefit of the doubt and provides financial support for an adjustment program. But the exposure of the Union should be limited. If the program goes awry, a cut in the debt must be considered dispassionately. The EMF could be of great help even if this has become unavoidable, as it could provide bridge financing and a framework for negotiations between the creditors and the debtor country.

APPENDIX: A STANDSTILL FOR DEBT SERVICE DURING THE ADJUSTMENT PROGRAM

It appears very difficult to limit ESM financing for a country with high public debt when the country's debt has not been judged unsustainable at the beginning of a program. Consider the example of a country with a debt-to-GDP ratio close to 140% and an average maturity of seven years. With this combination, about 60% of GDP would have to be refinanced over the first three years of a potential ESM program. To this burden one would almost surely have to add some current deficits, which over three years could easily add another 10% of GDP. The initial program could thus require 70% of GDP. But this would lead to a situation in which there would be little room for any haircut, if the program did not succeed in restoring growth and hence external and fiscal balance. Very short-term debt, which is almost never subject to a haircut, typically amounts to about 15% of total debt, and would in this case likely be worth 20% of GDP.

The banks of the country concerned might hold another 20% worth of GDP in bonds. This debt could not be cut either, because that would destroy the financial system of the country (and any chance of the program succeeding).

On top of this, one would have to consider any holdings of home country public debt by the national central bank. The bonds held by the central bank (acquired, for example, under the Public Sector Purchase Program of the ECB) might be formally subject to a haircut (as long as the national central bank does not hold a blocking minority). But this does not help because the national central bank is part of the national public sector and any losses it incurs would fall back on the government anyway.

REFERENCES

Brunnermeier, M. K., H. James, and J.-P. Landau, 2016, The Euro and the battle of ideas, Princeton University Press

De Groen, W. P., and D. Gros, 2015, "Estimating the bridge financing needs of the Single Resolution Fund: how expensive is it to resolve a bank?" in-depth analysis requested by the Economic and Monetary Affairs Committee of the European Parliament, PE 542.687, DG for Internal Policies, Brussels, November, https://bit.ly/2OcPEOB

Farhi, E., and J. Tirole, 2017, "Deadly embrace - sovereign and financial balance sheets doom loops," EconPol Working Paper 1/2017, European Network for Economic and Fiscal Policy Research, https://bit.ly/2zxuFh1

Gros, D., and T. Mayer, 2010a, "Disciplinary measures," The Economist Economic Focus, https://econ.st/2Q5S3Yo

Gros, D., and T. Mayer, 2010b, "How to deal with sovereign default in Europe: towards a Euro(pean) Monetary Fund," Report requested by the European Parliament, https://bit.ly/2xBODWL

Gros, D., and T. Mayer, 2017, "A European Monetary Fund: why and how?" CEPS working document no. 2017/11, CEPS, December, https://bit.ly/2Bs5ApC

Schäuble, W., 2017, "Paving the way towards a Stability Union," non-paper presented to the ECOFIN meeting, October (see article in the Financial Times, 2017, "Paris and Berlin at odds over key plank in eurozone reform plans," 10 October, https://on.ft.com/2fXgpDV)

Tirole, J., 2015, "Country solidarity in sovereign crises," American Economic Review 105:8, 2333–2363

Table A1: Access limits (based on a multiple of actual ESM quotas)

	ESM QUOTA (€ BLN)	€ BLN	% GDP	ACTUAL TOTAL FINANCING RECEIVED AS A % OF THE ACCESS LIMIT
Germany	134.8	404.4	12.9	–
France	101.2	303.7	13.6	–
Italy	89.0	266.9	15.9	–
Spain	59.1	177.3	15.8	23.3
Netherlands	28.4	142.0	20.2	–
Belgium	17.3	86.3	20.4	–
Austria	13.8	69.1	19.8	–
Ireland	7.9	39.5	14.4	170.7
Finland	8.9	44.6	20.7	–
Portugal	12.5	62.3	33.7	125.2
Greece	14.0	69.9	39.7	371.8
Slovakia	4.1	28.6	35.4	–
Luxembourg	1.2	8.7	16.4	–
Slovenia	2.1	14.9	37.2	–
Lithuania	2.0	14.2	36.5	–
Latvia	1.4	9.6	38.4	–
Estonia	0.9	6.5	30.8	–
Cyprus	1.0	6.8	35.9	92.4
Malta	0.4	2.5	25.7	–

Note: Total financing includes IMF assistance (and for Ireland third-country contributions).
Source: Own calculations based on AMECO data and the ESM

The remaining debt subject to a haircut could consequently be reduced to only about 20-30% of GDP after three years of a program. The general corollary to these illustrative calculations is that there might be little room left if a highly indebted country is fully refinanced even over only a few years.

We witnessed something along these lines in the case of Greece, where officially €200 billion in debt instruments could be made subject to the private-sector involvement (PSI) operation. Very short-term debt (and that held by the ECB) was excluded. But the holdings of the Greek banks (and insurance companies) comprised a large part of the €200 billion, and they had to be refinanced by the ESM in order to keep the Greek banking system afloat. Only about €100 billion of bonds still held by international investors were effectively subject to the PSI, which reduced them in nominal terms to about €50 billion, resulting in a gain of about 25% of GDP. This outcome is actually somewhat better than what one would expect in the example given above: the country would start with a debt ratio of around 140%. If 30% remained available for cutting, this would yield a potential gain (in nominal terms) of about 15% of GDP, or a new debt ratio of 125% of GDP.

All these calculations were in terms of the initial GDP. But a country going into a deep crisis will typically experience, at least initially, a fall in GDP. Moreover, countries needing external financing tend to have lost competitiveness and will need an internal devaluation. This implies that, provided that the rest of the euro area sticks to price stability, nominal GDP might have to fall considerably before growth resumes. If nominal GDP must fall by 20%, the debt ratio at the end of the adjustment process would be 125 × 1.2 or 150% of GDP.

These considerations suggest several policy conclusions:

- There would have to be a standstill for debt service payments when a program is launched for a country with very high initial debt (this was not the case for Spain or Ireland).
- Countries with high debt ratios should be induced to have longer average maturity. Government debt management is considered a purely national prerogative. Yet in a crisis, the structure of public debt becomes a key issue for the entire area. The incentives, therefore, are not properly aligned: in the run-up to a crisis, the country usually starts issuing shorter-term debt because it wants to avoid locking in high-risk premia for a long time. If everything goes well, this will have been the right choice. But if the crisis deepens and the ESM has to intervene, the risk will be transferred to the euro-area taxpayer.
- Banks should not be allowed to hold large amounts of the debt of their own government. This prohibition is in the interest of their country, as substantial holdings of government debt could lead to a situation forcing the country to exit EMU in the case of a restructuring of government debt.
- Finally, more thought should be given to how to measure public debt ratios. Any overvalued domestic price levels should be taken into account in assessments of debt sustainability. Accordingly, this contribution proposes to limit access to ESM funding, as described in the main text. Table A1 provides an illustration of the resulting access limits.

REGULATION OF CROWDFUNDING

TOBIAS H. TRÖGER | Professor of Private Law, Trade and Business Law, Jurisprudence, Goethe University Frankfurt am Main, Program Director Research Center Sustainable Architecture for Finance in Europe (SAFE)*

ABSTRACT

This paper is a shorter version of the national report for Germany prepared for the 20th General Congress of the International Academy of Comparative Law 2018. It gives an overview of the regulation of crowdfunding in Germany and the typical design of crowdfunding campaigns under this legal framework. After a brief survey of market data, it delineates the classification of crowdfunding transactions in German contract and corporate law and their treatment under the applicable conflict of laws regime. It then turns to the relevant rules in prudential banking regulation and capital market law. It highlights disclosure requirements that flow from both contractual obligations of the initiators of campaigns vis-à-vis contributors and securities regulation (prospectus regime).

1. INTRODUCTION

1.1 Policy objectives

Crowdfunding is a buzzword that signifies a subset of the new forms of finance facilitated by advances in information technology, usually categorized as fintech.[1] In contrast to financial innovation that pertains to (new or redesigned) financial products and is somewhat ambiguous in terms of its social value,[2] crowdfunding capitalizes on previously unavailable digital techniques to match supply and demand on money and capital markets. These developments can potentially disrupt traditional forms of intermediation by shifting the boundaries of the (financial) firm.[3] Put differently, crowdfunding does not typically lead to unprecedented forms of financing relations. Instead, it allows for traditional contractual or corporate law relationships between previously unacquainted providers and consumers of capital to be initiated and concluded on novel, IT-driven platforms. From this perspective, the potential of crowdfunding to garner economically significant volumes of financing relationships seems considerable,[4] thereby creating massive potential for momentous disruption as a consequence of disintermediation.

Once these projected developments gain traction, policy objectives traditionally pursued in financial regulation also become relevant for agents involved in crowdfunding.[5] Concerns about financial stability, investor and consumer protection, or the prevention of money laundering and funding of terrorism hinge incrementally on including these new techniques to initiate financing relationships adequately in the regulatory framework. More specifically, the legislation through which policymakers seek to implement the relevant objectives, ceteris paribus, have to be attentive to the specifics of crowdfunding.

Considering the aforesaid, the pertinent legislation must pay particular attention to the role of the platforms and their operators because they are at the heart of the

* A longer version of this paper was published in "German National Reports on the 20th International Congress of Comparative Law" 397-428 (Martin Schmidt-Kessel, ed., Tübingen: Mohr Siebeck, 2018).

technological innovation, which may both attenuate traditional justifications for government intervention and create new jeopardies for established policy goals. On the other hand, the laws that govern the relevant financing relationships once they are concluded face far fewer challenges insofar as they are not materially affected by the way relationships are initiated and concluded. Put differently, the contract or corporate law framework that underpins financing relationships is old-fashioned, but the way it is invoked is novel.

1.2 Economic relevance of crowdfunding in Germany

The available data largely pertains to the forms of crowdfunding that initiate classical financing relationships (loan contracts, purchase of debt instruments, or equity interests). Granular data on funding relationships with significant altruistic elements is largely lacking.[6]

1.2.1 CROWDLENDING/PEER TO PEER (P2P) LENDING

In a study commissioned by the Federal Ministry of Finance, financial economists produced data inter alia on the scope and structure of the crowdlending market over the period from 2007 to 2015.[7] The findings showed an enormous growth of what is the largest segment of the crowdfunding market (totaling €400 million of credit extended by the end of 2015, with average annual growth rates of 95%)[8] with a significant slowdown during the economic downturn and even a decline of 22% in 2011. While P2P lending to consumers occurred relatively early on, crowdlending to businesses is a comparatively new phenomenon, albeit with staggering growth rates.[9] Until the end of the observation period, the market was dominated by one player (Auxmoney), mainly used to roll-over existing loans or overdrafts and exhibiting relatively high default rates.[10] This arguably induced platforms to impose stricter access conditions for users seeking credit (presentation of credit ratings). They thus assumed a more important role as gatekeepers.[11]

1.2.2 CROWDINVESTING

Germany's preeminent scholars in the field produced descriptive statistics on the domestic crowdinvesting market.[12] They showed not only that the initial upward trend in the funds raised (a total of almost €53 million since the first crowdinvestment initiative in August 2011) has abated recently,[13] but that fundraising is largely concentrated on two platforms (Seedmatch and Companisto). These key players are also highly successful in placing the issues of start-ups (the success rate was 100% and 95% respectively), whereas other platforms have a significant fraction of failed offers that do not reach the funding threshold. With all due reservations concerning methodologically unhedged inferences, the data seems to indicate that platforms perform gate-keeping functions[14] and are in a position to build reputational capital as information intermediaries as well.

2. DEFINITION – LEGAL QUALIFICATIONS OF CROWDFUNDING

2.1 How is crowdfunding defined in your legal order?

German law does not have any statutory or otherwise authoritative definition of crowdfunding. Scholars define crowdfunding as "collecting financial contributions from a multitude of persons to achieve a common goal through the use of a specialized internet platform."[15] Even more broadly, the German Federal Financial Supervisory Authority (Bundesanstalt für Finanzdienstleistungsaufsicht, BaFin), understands crowdfunding as "a type of financing which is usually raised over internet platforms."[16] Although definitions vary in detail,[17] the common recurring theme is that crowdfunding campaigns are conducted and supply and demand are matched over the Internet or through social media.

2.2 Situations usually covered by the notion 'crowdfunding'

Variations in the terminology of the German scholarly debate aside, it is useful to distinguish between several sub-categories of crowdfunding. They are characterized by the diverging objectives that parties pursue with their transactions, which in turn shape the considerations stipulated in the contract.[18] In crowdsponsoring, contributors receive no financial compensation, but support a specific project with donations.[19] Alternatively, contributions are rewarded with (nominal) non-monetary benefits ("goodies") if the campaign is successful, like an acknowledgement on the cover of music media or in the credits at the end of a movie.[20] Alternatively, the consideration can have material value, for instance if supporters of crowdfunding campaigns receive a product from the first batch of production or acquire the preferential right to purchase the product immediately at a reduced

price (reward-based or pre-selling crowdfunding).[21] The funding relationship exhibits an even clearer character as an exchange agreement if a financial consideration is stipulated,[22] either as fixed compensation (interest) for the temporary provision of liquidity[23] (crowdlending) or as variable, performance-related payment that flows from investments in a business venture in the form of equity or mezzanine-capital instruments (crowdinvesting or commercial crowdfunding).[24] Finally, a similar arrangement occurs where supporters participate in the exploitation of copyrights, patents, and similar intellectual property rights that were facilitated through their crowdfunding contributions, for instance by receiving a share of the royalties paid to an artist.[25]

2.3 Legal qualifications for the different types of funding

The general stance of German contract law towards crowdfunding is determined by the fundamental principle of freedom of contract.[26] This holds true even for crowdinvesting instruments that grant sponsors participation rights in a business venture's future cash flows, because, as a matter of law, the hybrid capital instruments typically offered constitute debt contracts that are unaffected by corporate law's rigidity.[27] This latitude enables initiators of crowdfunding campaigns to structure the respective financing relationships to fit their preferences. Yet, it should not be ignored that the latter are frequently shaped by an appetite to avoid the constraints of banking and securities regulation. However, as initially noted, the legal qualification of financing relationships concluded on platforms poses no idiosyncratic challenge for German private law, because, in principle, all funding relationships existed prior to digitization in the analogue world and technological innovation has only facilitated their conclusion among previously unacquainted parties.

2.4 Crowdsponsoring

If contributions to the campaign are made as donations or no-interest "loans" without repayment-obligation the qualification as an immediately executed gift contract (Handschenkung) within the meaning of § 516 BGB is straightforward.[28] The classification requires that the contribution is made without consideration, meaning the grant does not legally depend on any return, however small.[29] Quite importantly, promises of non-monetary rewards also qualify as a consideration that precludes the qualification of a contract as a gift contract.[30] However, crowdfunding campaigns where initiators promise no more than to publicly announce the name of the contributor do not necessarily provide for such non-monetary compensation. If the mentioned name is only one among many others of those who made (small) contributions, the typical credits can be qualified as legally irrelevant references to the gift.[31] Only if the contribution that is supposed to be mentioned is more prominent, and thus allows for increased (media) attention can the relationship between the initiator and the contributor qualify as a sponsoring contract.[32] In these contracts, the publicity of the contribution materially serves the communicative purposes of the benefactor and its promise thus constitutes a relevant compensation for the granted funds.[33]

Moreover, German private law requires that both parties agree that the contribution occurs without consideration. Simply put, there must be contractual consensus on its gratuitousness.[34] Such a consensus exists when the contribution is neither in a synallagma with a consideration, nor the condition, nor the cause of law for such a quid pro quo.[35] Hence, if contributors enter into a legally binding arrangement promising them a material advantage in the form of an incentive or a goody (for instance a free download of funded music productions or meeting with the artist), the contract cannot be comprehensively qualified as a donation.[36] However, if the parties are aware of a significant mismatch between the higher value of the contribution and the lower one of the consideration, German doctrine splits the transaction into two independent contracts,[37] and thus treats the overshooting fraction of the contribution as a donation[38] and treats its compensated part as a reward-based crowdfunding contract.[39]

2.5 Reward-based crowdfunding

If investors in successful crowdfunding campaigns receive access to the product as a consideration for their contribution, for instance a physical delivery from the first manufacturing batch, a data medium with the produced movie or music album or a download code for it, the underlying contract can easily be qualified as a sale.[40] If contributors acquire only a right to buy the product (at a reduced price), the contractual relationship is a purchase of rights, which is explicitly qualified as a sale in BGB § 453 para. 1.[41] If media can only be streamed and no download-to-own is possible, the contractual relationship between investors and benefactors of crowdfunding campaigns represents a rental agreement.[42] Generally, if the product value (market price) or the price of the

acquired right is – in accordance with the parties' agreement – lower than the contribution, the transaction may be treated as consisting of two separate contracts.[43]

2.6 Crowdlending/P2P lending

P2P lending leads to regular, typically unsecured loan agreements.[44] Loans to finance the acquisition of real estate, in principle subject to the same provisions in the German civil code, are practically non-existent, because such transactions are typically executed through special purpose vehicles in crowdinvesting (infra 4).

However, direct contracting between lenders and borrowers, mediated through the platform as an agent, would trigger undesirable regulatory consequences[45] and is, therefore, rare in Germany, as operators have adjusted their business models accordingly. Although platforms match lenders and borrowers, they interpose a credit institution in the transaction that contracts with both the credit-seeking party and the funding party.[46] On the one hand, the borrower takes out a loan from the credit institution, procured by the platform that earns a service fee (borrowing fee). On the other hand, the funding party purchases the bank's redemption claim, which is subsequently assigned once the bank disburses the loan.[47] As an economic result of the transaction, the investor holds a claim against the borrower just like they would had they contracted directly.[48] This observation begs the question whether a differential treatment in regulation (see infra D.I.1) can be justified as a matter of public policy.

2.7 Crowdinvesting

Contributors to crowdinvesting campaigns receive a variable compensation that hinges on the financed venture's future cash flows. The specific design of the arrangements varies[49] and the observable differences are relevant for the legal qualification of the contractual relationships the parties typically conclude. In the vast majority of cases, the project-executing organization or person enters into direct contractual relationships with investors through the platform, whereas arrangements in which a special purpose entity bundles investments and then contracts with the initiator are rare.[50]

Recent empirical research highlights the legal structure of typical crowdinvestment products offered through the platforms to finance business ventures.[51] These insights are of critical importance, because they determine how and to what extent crowdinvesting affects the policy objectives of financial regulation. The legal structure of investment products sold on crowdinvesting platforms defines both the cash-flow and governance rights vested with investors, which in turn are crucial for investor protection, but also have an impact on financial stability.

Issuers typically structure the financing relationship as unsecuritized term-debt[52] with fixed interest rates[53] and various extents of profit participation.[54] In most cases, investors also participate in an increase of the going-concern value of the issuer.[55] Loss participation is limited to the funds invested in gone-concern scenarios.[56] Contractual arrangements in the indenture subordinate

the redemption claim to all other claims against the issuer.[57] The contractual relations that underlie typical German crowdinvestments seek to mimic equity-like risk-and-return structures. This becomes even more apparent when considering the protection against claim dilution in the case of follow-up funding,[58] which prevents new investors from externalizing risk to old investors and benefiting disproportionately from future cashflows.

However, the governance rights granted to investors on crowdinvesting platforms are limited compared to those vested with shareholders. In essence, investors do not have any influence on the decision-making process of the issuer concerning questions of management and business strategy.[59] Yet, contracts provide for periodic disclosure of key financial and other relevant data that in some cases have to be explained by initiators at web-based annual investor meetings.[60] Control rights beyond the entitlement to candid disclosure are almost non-existent.[61]

In essence, German law provides three types of contractual arrangements that conform to the rights and obligations the parties seek to establish in crowdinvesting transactions.[62] The relationship between contributors and initiators of crowdinvesting campaigns can be framed as either silent partnerships,[63] profit participation rights (Genussrechte),[64] or subordinated profit-participating loans (partiarische Nachrangdarlehen).[65] The precise classification of individual agreements is difficult and courts explicitly follow a case-by-case approach.[66] However, key indicators are (i) the lack of monitoring and control rights, which militates against a qualification as (silent) partnership;[67] (ii) the existence of a fixed repayment claim combined with a participation in the venture's profits or turnover, which speaks in favor of a profit participating loan contract;[68] and (iii) the absence of such a repayment claim and a loss participation not only in gone-concern scenarios that hints at the classification of the financing relationship as a profit participation right or a silent partnership.[69] To distinguish between profit participation rights and silent partnership interests, a pivotal factor is whether the crowdfunding relationship obliges contributors to further the project (common purpose) beyond their financial contribution.[70]

At times, commentators have sought to establish a separate category for single-project financing relationships like movie productions or music albums.[71] However, this further distinction is unnecessary, as these contracts can be understood as loans with (subordinated) fixed repayment obligations,[72] and the value of the latter hinges on the performance of a single asset and thereby leads to an automatic loss-participation of investors up to the contributed amount. Alternatively, the respective financing relationships can also be construed as profit participation rights granted by the producing entity, where no repayment claims exist and a loss-participation is possible.[73]

3. NORMATIVE FRAMEWORK

3.1 General

There has been no legislative intervention with regards to the private law qualification of contracts concluded on crowdfunding platforms, probably because the existing German private law framework allows parties to structure their financing relationships according to their economic goals. They can draw on well-established and thus broadly approved doctrinal concepts, which are applied to crowdfunding activities.[74] Deviations from the majority view in the literature are confined to narrow aspects, remain exceptions, and are ultimately not convincing.[75]

There is no specific law that regulates crowdfunding. Only very limited legislative interventions exist that relax primary market disclosure obligations in securities laws for crowdfunding activities.[76]

3.2 Conflict of laws

Typical financing relationships concluded on platforms (see supra B.III) fall within the remit of the Rome I Regulation.[77] This is also true for the most common crowdinvesting contracts, the subordinated profit participating loans (which are not negotiable instruments within the meaning of art. 1 para. 1 lit. d) Rome I Regulation[78]), unsecuritized profit participation rights, and – according to the majority view in the literature – silent partnership interests.[79] Although company law relationships are generally exempt from the regulation's scope of application,[80] silent partnerships, by their very nature, do not entail an actual organization but establish only contractual ties between the partners.

For all prevalent forms of crowdfunding, a choice of law is thus possible in principle.[81] There is no publicly available empirical evidence on whether the option is broadly used in practice.[82] In any case, the European conflict of laws rules limit the possibility to choose the applicable law in consumer contracts insofar as the

consumers would be deprived of the protection afforded to them by provisions that cannot be derogated from agreement by virtue of the law of the country of the consumer's habitual residence.[83] This rule applies in crowdfunding relationships concluded through German platforms, because even in crowdinvestment the relevant contracts do not establish rights and obligations that constitute a financial instrument within the meaning of the exception from the binding consumer protection afforded under the Rome I Regulation.[84] However, where platforms seek to derogate from German law, the most important consumer protection rules to be considered in the required comparison with the chosen legal system are the subscription limits stipulated in securities laws.[85]

> "The pertinent legislation must pay particular attention to the role of the platforms and their operators because they are at the heart of the technological innovation, which may both attenuate traditional justifications for government intervention and create new jeopardies for established policy goals."

Where choice of law clauses is not introduced in the respective contracts and consumer protection rules do not apply,[86] the relationship is governed by the law of the country where the party required to effect the characteristic performance has their habitual residence.[87] In crowdfunding relationships, this means the law of the country where the contributor lives.[88]

4. SUPERVISION OF CROWDFUNDING ACTIVITY

4.1 Licensing requirements

Germany has no specific prudential regulation for crowdfunding. Authorization requirements can, therefore, only flow from the general bodies of law that regulate the financial sector, in particular the regulations governing credit institutions and investment services firms. The intermediation of donation- and reward-based crowdfunding does not constitute an activity that can fall under the regimes of prudential banking and capital market regulation, as long as platforms avoid collecting the funds from contributors beforehand.[89] However, the cases of crowdlending (infra D.I.1) and crowdinvesting (infra D.I.2) are less straightforward and largely depend on platforms' business models. Where the business model leads to licensing requirements, the applicable regime for obtaining and withdrawing licenses is that which is prescribed for credit institutions and investment firms respectively (infra D.I.3).

4.2 Crowdlending/P2P lending

Whether crowdlending platforms require an authorization under the Banking Act hinges on whether their activity is classified as either banking business or financial service.[90]

From the outset, there is a broad consensus that the primary economic function of platforms, to broker credit, does not constitute banking business within the meaning of the law,[91] particularly because simple loans do not represent financial instruments and hence the activity of platforms does not amount to investment brokerage (Anlagevermittlung).[92] Yet, within this function, specific intermediate steps may amount to banking business and thus trigger the authorization requirement.

If platforms collected the monetary contributions from the crowdlenders before forwarding them to borrowers, they might fulfil the statutory elements of "deposit business" (Einlagengeschäft).[93] Although even registered users of the platform would provide "public funds" as required by the law,[94] platforms can avoid falling under prudential banking regulation by not offering lenders accounts, and collecting the funds in successful campaigns only after the threshold level has been reached and forwarding them as quickly as technically possible to borrowers. This already avoids the funds being regarded as being "taken" by the platform.[95] Platforms are even safer if they have contributions collected and forwarded by a cooperating bank, thereby avoiding the acceptance of lenders' funds in the first place.

The challenges faced by platforms when they wish to avoid their activities amounting to "credit business" (Kreditgeschäft)[96] are far more daunting. As long as platforms do not issue credits themselves, they do not violate a pre-authorization requirement with their own conduct.[97] However, they may be held liable for aiding and abetting others in such an infringement of the banking monopoly and supervisors may, therefore, enjoin their

operations.[98] Contributors themselves may fall under the very extensive interpretation of "credit business" and, therefore, conduct unauthorized banking operations.[99] Any person that extends money loans engages in "credit business" if the activity is commercial.[100] According to the majority view endorsed by supervisory practice, an activity is commercial if it is intended for a certain time period and motivated by an intent to achieve profits.[101] A single transaction may suffice, if the intention is to extend more loans in the future.[102]

Platforms react to the extensive authorization requirement by favoring the indirect contracting model (supra B.III.3).[103] Despite economically identical outcomes, the supervisory practice and the majority view in the literature accept that combining the transactions does not amount to "credit business" for any other party involved than the loan-originating bank, and this can, therefore, be conducted without (additional) banking licenses.[104] In particular, the various activities of platforms in the indirect contracting models also do not constitute banking business.[105]

4.3 Crowdinvesting

Licensing requirements for crowdinvesting platforms[106] under the Banking Act hinge on whether their activity qualifies as either banking business or an investment service.[107]

Regardless of the statutory stipulations of specific activities in the statutory definitions, any financial and investment service has to pertain to "financial instruments" as defined in banking and securities regulation.[108] Prior to June 1, 2012, silent partnership interests and unsecuritized participation rights were not included in this definition, essentially liberating crowdinvesting platforms from any authorization requirement and the prudential supervision attached to it. Since then, the definition of financial instruments also encompasses "financial assets" within the meaning of the Capital Investment Act,[109] and since July 10, 2015, these in turn also comprise subordinated profit participating loans.[110] Hence, the regulatory framework now in principle also captures the typical OTC investment products offered through platforms, like silent partnership interests, participation rights, or subordinated profit participating loans.

Consequently, the main query has become whether the activity of crowdinvesting platforms with regard to financial instruments constitutes one of the enumerated business activities that qualify as banking or investment services. The consensus among scholars is that platforms do not engage in underwriting business (Emissionsgeschäft),[111] because they do not assume the risk of a successful placement of the financial instruments issued.[112] Similarly, typical platform activities do not constitute placement business (Platzierungsgeschäft),[113] because this would require that the platform acts as an agent of the issuer and – according to the interpretation of BaFin – discloses this agency relationship.[114] Instead, platforms typically only deliver offers to buy or sell as messengers.[115] However, despite some quibbles about the precise meaning of the law,[116] platforms may indeed engage in investment brokerage (Anlagevermittlung),[117] because they intermediate the acquisition and sale of financial instruments.[118] According to the majority view, it does not matter whether the transactions occur on the primary or secondary market.[119] Hence, the execution of initial offerings through crowdinvesting platforms may fall under the definition of investment brokerage and thus constitute banking or investment services that, in principle, require authorization. Nevertheless, brokerage activities that pertain to financial assets are exempt from authorization requirements if brokers acquire property rights neither in the assets nor in the invested funds of the customers.[120] This tallies perfectly with the typical business model of crowdinvesting platforms. As a consequence, only a special form of trade supervision (qualifizierte Gewerbeaufsicht) applies.[121]

Finally, authorization requirements could be attached if a platform's activities constitute the operation of a multilateral trading facility (MTF).[122] Some commentators unconvincingly rule out this possibility by pointing to the regulatory rationale of the underlying European legislative initiatives that sought to capture MTFs as contemporary competitors of exchanges, arguing that this would require that platforms also host secondary market trading.[123] The relevant policy goal of the pertinent regulation is to counter efficiency losses that are associated with a fragmentation of trading. In this regard, price discovery on primary markets is just as important as it is on secondary markets.[124] The German supervisor has also repeatedly published the interpretation that crowdinvesting platforms can fall under the definition of MTFs.[125] However, it is unclear under which preconditions BaFin will actually find that the specific requirement of a "large number" of market participants trading at an MTF has been met in crowdinvestment initiatives.[126]

4.4 Licensing regime

If German crowdfunding platforms chose business models that require an authorization as a credit institution or an investment firm, they would have to fulfil all the requirements put forward in prudential banking or securities regulation, in particular the own funds requirements applicable to banks[127] and the extensive standards for the conduct and the organization of financial services firms.[128] Failure to comply would lead to licenses being revoked by the European Central Bank (banking license)[129] or BaFin (financial services firms).

5. SPECIFIC OBLIGATIONS OF PARTIES

5.1 Disclosure requirements

Obligations beyond regular contract law only apply to crowdlending and crowdinvesting.

5.1.1 CROWDLENDING

As a consequence of the indirect contracting model, the bank that cooperates with the platform has to fulfil the extensive disclosure obligations stipulated for consumer loans,[130] as prescribed in European law.[131] The platform itself incurs a duty to disclose information on the specifics of its involvement and the remuneration received for it.[132]

5.1.2 CROWDINVESTING

Funding an unseasoned business without a robust track-record is fraught with informational asymmetries between investors and founders (insiders) that typically lead to adverse selection problems.[133] These are all the more serious in our context, because the likelihood of failure of a funded venture and thus a default on investors' claims is usually high in crowdinvesting.[134] As a consequence, information obligations vis-à-vis investors are pivotal. These can follow either from contractual obligations to inform (infra A.I.1.a) or the prospectus requirement put forward in securities regulation (infra A.I.1.b).

5.1.2.1 Contractual obligation of platforms

Although platforms typically do not perform the role of an investment advisor with the respective set of extensive duties[135] simply because they do not recommend specific investments,[136] some commentators argue that they incur contractual obligations to provide specific information to investors as an investment broker.[137] The main argument is that, by pre-screening investments and structuring information presented to the crowd, platforms solicit trust in their superior expertise and access to information that investors rely upon.[138] However, others hold that platforms advertise investments without an intent to incur legally binding information obligations.[139] The latter position is not convincing given German courts' general tendency to generously presume tacit agreements where information asymmetries are striking.[140] Moreover, the practice of platforms not to gather, assess, and provide information is irrelevant with regard to establishing potential obligations and potentially amounts to neglectful behavior.

According to general standards, platforms, therefore, have an obligation to fully and correctly provide all information they possess that is material for the investment decision to be made.[141] Furthermore, they have to verify the plausibility of the information supplied by the initiator of the campaign.[142] This means, as a minimum, they have to assess whether the initiator provided all material information investors need to gauge the risks inherent in the investment (for instance on the project idea, business plan, specific risks, management, legal form of business venture, and investment) and to disclose information gaps, if the initiator's submission proves insufficient and additional data is unavailable.[143] Some commentators argue that platforms additionally have to roughly evaluate the viability of the venture, in order to weed-out "evidently extreme examples" of unrealistic business models.[144]

5.1.2.2 Prospectus requirement and investor information sheet

An important potential channel through which information asymmetries between issuers and investors can be countered in crowdinvesting are prospectus requirements. As intermediaries, platforms cannot have an original duty to draw-up a registration document themselves, but can serve as powerful gatekeepers, if the general prohibition to distribute financial instruments without a prospectus[145] also applies for investments initiated and concluded through crowdinvesting platforms.

Until July 10, 2015, a full-blown prospectus requirement under VermAnlG, § 6 for offerings with a nominal value of more than €100,000 existed, yet certain financing relationships, in particular subordinated profit participating loans, were generally not captured by the regime.[146] The reform package of the Small Investor Protection Act[147] closed the loopholes, but established

an exemption for financial assets offered through crowdinvesting platforms (Schwarmfinanzierung).[148] The main preconditions[149] are that the aggregate value of the offering does not exceed €2,500,000, that subscription limits that depend on net worth and income of investors range from €1,000-€10,000,[150] and that compliance with these preconditions is monitored by the platform. The primary source of information becomes the mandatory investment information sheet (Vermögensanlagen-Informationsblatt), which must be prepared by issuers and provided to potential investors who have to confirm that they (read and) understand a specific warning that points to the risk of a total loss of the invested funds.[151] It has to contain an explicit notice that no prospectus was prepared for the offering.[152] The advertisement restrictions, to be enforced by BaFin,[153] ensure that the express warnings prescribed by law do not go missing in any other relevant communication regarding the investment.[154]

Issuers on crowdinvesting platforms thus have limited choice regarding the regime for primary market disclosure.[155] They can either opt for a fully-fledged prospectus and offer their product publicly without restrictions or accept limitations and make use of the statutory exemption provided for crowdinvesting.

5.2 No obligation to guarantee accomplishment of, or follow-up on, the project?

German law does not provide for an obligation to guarantee the accomplishment of the project or the participation in follow-up projects. However, typical contractual arrangements contain all-or-nothing clauses that ensure that initiators will only draw on individual contributions if the campaign reaches the target volume of financing.[156] Hence, contributors have at least some certainty that the preconditions for successfully initiating the project are met. Moreover, some protections against abusive practices ex-post exist, most importantly the obligation to pay damages if the initiator misappropriates the funds received.[157]

5.3 Redress mechanisms in case of non-accomplishment of the project

If a crowdfunding project fails due to the breach of a specific contractual obligation and there is a finding of fault on the side of the party in breach, damages may be available.[158]

Platforms can only be liable for a breach of an obligation to inform. Such duties are most prominent in crowdinvesting where platforms may assume a role as investment brokers subject to specific information obligations (supra A.I.1.a), with a rich body of case law substantiating the respective duties.[159]

Fraudulent behavior aside, project directors may be liable if they deploy funds in a way that contradicts the project description in the campaign. This can occur through a breach of the primary obligation to produce a certain good (reward-based crowdfunding) or violate the secondary obligation to avoid any action that imperils the other party's contractual objectives (crowdsponsoring, crowdinvesting).[160] Whether a deviation from the original plans was a good faith attempt to achieve the original goals of the campaign or a misappropriation of funds is often difficult to discern.

In principle, unsound managerial decisions that are not in line with acceptable business practice can give rise to liability.[161] However, although no specific case law is available, courts will probably be reluctant to find fault in business decisions, as long as they were made on a sound informational basis and in the absence of conflicts of interest.[162]

ENDNOTES

[1] On fintech in particular, Zetzsche, D. A., R. P. Buckley, D. W. Arner, and J. N. Barberis, Forthcoming, "From FinTech to TechFin: the regulatory challenges of data-driven finance," New York University Journal of Law and Business.

[2] For proposals that seek to hedge financial stability against regulatory arbitrage without sacrificing the efficiency enhancing potential of financial innovation see Posner, E., and E. G. Weyl, 2013, "An FDA for financial innovation: applying the insurable interest doctrine to 21st century financial markets," 107 Northwester Universty Law Review, 1307 (arguing for pre-screening of financial innovations through a Federal Drug Authority like agency); Tobias H. Tröger, T. H., 2016, "How special are they? Targeting systemic risk by regulating shadow banks," in Lomfeld, B., A. Somma, and P. Zumbansen (eds.), Reshaping markets: Economic governance and liberal utopia, Cambridge University Press, 185-207 (showing how a normative approach to law enforcement allowed existing prudential regulation to capture regulatory arbitrage). For an overview of the regulatory challenges non-bank banks pose with regard to systemic risk, Wymeersch, E., 2017, "Shadow banking and systemic risk," EBI working paper no. 1.

[3] The extent to which allocation of resources occurs in a hierarchy (firm) depends on the transaction costs incurred in equivalent market transactions, for the fundamental insight, Coase, R. H., 1937, "The nature of the firm," 4 Economica 386; for a review of the literature carrying forward the theory of the firm see Furubotn, E. G., and R. Richter, 2005, Institutions and economic theory, University of Michigan Press, 2nd edition, 366-386. With regard to financial intermediation, this means that market-based solutions should become more prominent once the comparative advantages of intermediation within a big entity shrink, which is particularly the case if search costs are lowered as a function of technological improvements.

[4] The early literature points to crowdinvesting's potential to allow firms to receive financing from an additional source that complements bank and venture capital funding, Pope, N. D., 2011, "Crowdfunding microstartups: it's time for the Securities and Exchange Commission to approve a small offering exemption," 13 University of Pennsylvania Journal of Business Law, 101, 113; Hemingway, M., and S. R. Hoffman, 2011, "Proceed at your peril: crowdfunding and the Securities Act of 1933," 78 Tennessee. Law Review, 879, 931; Bradford, C. S., 2012, "Crowdfunding and the Federal Securities Laws," Columbia Business Law Review, 1, 103-104; for an account in the business press that envisions a far-reaching substitution of banks as providers of credit see Editorial, "Banking without banks," The Economist, 1 Mar 2014, at 70; for a delineation of crowdinvesting's potential in Europe see Zetzsche, D. A., and C. Preiner, 2017, "Cross-border crowdfunding – towards a single crowdfunding market for Europe, EBI working paper no. 8; for Germany, for instance, Meschkowski A., and F. K. Wilhelmi, 2013, "Investorenschutz im Crowdinvesting," 68 Betriebs-Berater (BB) 1411; specifically on the idea of a relative decrease in the costs of capital as a result of lower search and agency costs in crowdinvesting relationships, Klöhn, L., and L. Hornuf, 2012, "Crowdinvesting in Deutschland," 24 Zeitschrift für Bankrecht und Bankwirtschaft (ZBB) 237, 256-8 (arguing that the 'wisdom of crowds' is imperfect and partly irrelevant with regard to relevant agency relationships.

[5] For an overview of the policy issues, see Armour, J., and L. Enriques, 2017, "The promise and perils of crowdfunding: between corporate finance and consumer contracts, European Corporate Governance Institute working paper no. 366/2017; Zetzsche and Preiner (2017), supra note 4 at 9-16. The European Securities and Markets Authority (ESMA) has also identified what it considers to be key components for an adequate regulatory reaction to the new phenomenon and outlined several specific responses that draw-on and develop the existing E.U. regulatory framework, ESMA, Opinion: investment based crowdfunding 10-12 and 12-27.

[6] But see Dorfleitner, G., and L. Hornuf, 2016, "FinTech Markt in Deutschland," 22-25, https://bit.ly/2fwSa1C (presenting aggregate data for the donation- and reward-based crowdfunding markets that includes campaigns initiated by Germans on international platforms and showing that the overall funding capacity - €85 million between 2007 and 2015 – is small relative to other crowdfunding markets and dominated by three players, although sourcing occurs through a large number of intermediaries).

[7] Dorfleitner and Hornuf (2016) supra note 6 at 32-5. For older data, see Renner, M., 2014, "Banking without banks?" Rechtliche Rahmenbedingungen des Peer-to-Peer Lending, 26 ZBB 261, 262.

[8] This observation tallies with the global trend see Renner (2014) supra note 7 at 263.

[9] See also Zhang, B., R. Wardrop, T. Ziegler, A. Lui, J. Burton, A. James, and K. Garvey, 2016, "Sustaining momentum: The 2nd European Alternative Finance Industry Report," University of Cambridge study, https://bit.ly/2cMlUD9 (showing P2P-lending to consumers growing slower than P2P-lending to businesses, although on a higher level between 2013 and 2015). Projections indicate that the trend will continue in the future, see Statista, 2018, "Alternative lending segment report," https://bit.ly/2OqBlAE (reporting annual growth rates of 45.7% in the P2P business lending market until 2022, ultimately reaching €2,658 million).

[10] See also Dorfleitner, G., C. Priberny, S. Schuster, J. Stoiber, M. Weber, I. de Castro, and J. Kammler, 2016, "Description-text related soft information in peer-to-peer lending – evidence from two leading European platforms," 64 Journal of Banking and Finance 169 (reporting default rates of 12-14% for Germany's leading lending platforms).

[11] On the procedures of German crowdlending platforms see also Renner (2014) supra note 7 at 263.

[12] Klöhn, L., L. Hornuf and T. Schilling, 2016, "Crowdinvesting-Verträge," 28 ZBB 142, 143-5; for similar observations see Dorfleitner and Hornuf (2016) supra note 6 at 26-31; more recent data for 2016 corroborates the general trend, see Statista, 2018, "Gesamtvolumen des durch Crowdinvesting eingesammelten Kapitals in Deutschland von 2011 bis Q4 2016," https://bit.ly/2RaHJ2V. For additional empirical evidence see also Herr, S., and U. Bantleon, 2015, "Crowdinvesting als alternative Unternehmensfinanzierung – Grundlagen und Marktdaten in Deutschland," 53 Deutsches Steuerrecht (DStR) 532, 535 (2015); for granular data on the early phase, Klöhn and Hornuf (2012), supra note 4 at 239-46.

[13] See also Christopher D., 2016, "Crowdinvesting – Ist das Kleinanlegerschutzgesetz das junge Ende einer innovativen Finanzierungsform," 28 ZBB 20, 22 (observing above average growth of crowdinvesting only in real estate, ecological projects, and movie financing).

[14] For anecdotal evidence on very high rejection rates of up to 99%, see also Hornuf, L., and A. Schwienbacher, 2015, "The emergence of crowdinvesting in Europe: with an in-depth analysis of the German market," 25 note 12, LMU discussion paper 2014-43, https://bit.ly/1rqGpLq.

[15] Klöhn, L., L. Hornuf, and T. Schilling, 2016, "Regulation of crowdfunding in the German Small Investor Protection Act: content, consequences, critique, suggestions," 13 European Company Law 56 note 3; see also Klöhn and Hornuf (2012) supra note 4 at 239. This tallies with definitions present in the international literature, see, for instance, Hazen, T. L., 2012, "Crowdfunding or fraudfunding – social networks and the Securities Laws – why the specially tailored exemption must be conditioned on meaningful disclosure," 90 North Carolina Law Review 1735, 1736 (defining crowdfunding as sub-category of crowdsourcing "which refers to mass collaboration efforts through large numbers of people, generally using social media or the Internet"); similarly Hemingway and Hoffman (2011) supra note 4 at 881.

[16] https://bit.ly/2OZUAU1

[17] See, for instance, Meschkowski and Wilhelmi (2013) supra note 4 at 1411 (referring to the Wikipedia definition, which also highlights the collective effort in raising resources over the internet to support projects); see also Jansen, J. D., and T. Pfeifle, 2012, "Rechtliche Probleme des Crowdfunding," 33 Zeitschrift für Wirtschaftsrecht (ZIP) 1842, 1843 (pointing to the origins of crowdfunding in sponsoring charitable or altruistic projects through web-campaigns).

[18] For a distinction between altruistic and financially motivated crowdfunding see for instance, Sixt, E., 2017, Schwarmökonomie und Crowdfunding, Springer, 57. For a prudential supervisor's distinction of loan-based crowdfunding on the one hand and investment-based crowdfunding on the other, see Financial Conduct Authority, 2014, "The FCA's regulatory approach to crowdfunding over the internet, and the promotion of non-readily realisable securities by other media 5-6," Policy statement 14/4, https://bit.ly/2xSXL9A.

[19] A very effective example was Barack Obama's fundraising campaign for his initial Presidential Campaign, see Bradley, T., 2008, "Final fundraising figure: Obama's $750M," ABC News, December 5, https://abcn.ws/2RhAoOR.

[20] Jansen and Pfeifle (2012) supra note 17 at 1843; Bareiß, A., 2012, "Filmfinanzierung 2.0," Zeitschrift für Urheber- und Medienrecht (ZUM) 456, 460. For a delineation of possible designs, see also Schramm, D. M., and J. Carstens, 2014, "Startup-Crowdfunding und Crowdinvesting: Ein Guide für Gründer, Springer, 7.

[21] Sixt (2017) supra note 18 at 113.

[22] For a delineation of the respective categories see Bodensiek, K., and C. Leinemann, 2015, "Rechtliche Einordnung des Crowdfundings in Deutschland," 3-4 (Revue générale du droit, Études et réflexions No. 5, 2015), https://bit.ly/2lu7Rl9; Forster, M., 2013, Crowdfinancing 19 (2013); Sixt (2017) supra note 18 at 57-8; Schramm and Carstens (2014) supra note 20 at 7.

[23] On this fundamental feature of loan contracts, see Tröger, T., 2012, Loan, in Basedow, J., K. J. Hopt, and R. Zimmermann (eds.), The Max Planck Encyclopedia of European Private Law 1106, 1108-9.

[24] Specifically, on the definition of crowdinvesting as a form of financing of companies by granting an interest in the firm's future cash-flows, Tröger, T. H., 2017, "Remarks on the German regulation of crowdfunding," 12 Revue Trimestrielle de Droit Financier (RTDF) 79; similarly Klöhn et al. (2016) supra note 15 at 56 note 4 (2016); Klöhn and Hornuf (2012) supra note 4 at 239.

[25] Jansen and Pfeifle (2012) supra note 17 at 1843; Bareiß (2012) supra note 20 at 461-2.

[26] Bürgerliches Gesetzbuch [BGB] [Civil Code], Aug. 18, 1896, RGBl. at 195, § 311 para. 1 (translation at https://bit.ly/2QI7rQQ), empowers any person to create an obligation by contract.

[27] The German stock corporation law does not allow any material alteration of the statutory rights and duties of shareholders, Aktiengesetz [AktG] [Stock Corporation Act], Sept. 6, 1965, BGBl. I at 1089, § 23 para. 5. Even for other legal forms of business organizations, German law adheres to the principle of numerus clauses limiting the latitude to customize membership interests, see for instance Schmidt, K., 2002, Gesellschaftsrecht, Heymanns, 96-8 (4th ed.).

[28] Jansen and Pfeifle (2012) supra note 17 at 1843; Bareiß (2012) supra note 20 at 460.

[29] Reichsgericht [RG] [Imperial Court of Justice] Jan. 30, 1940, 125 Entscheidungen des Reichsgerichts in Zivilsachen [RGZ] 380 (383); Bundesgerichtshof [BGH] [Federal Court of Justice] Sept. 23, 1981, 82 Entscheidungen des Bundesgerichtshofs in Zivilsachen [BGHZ] 274 (280-2); Koch, J., 2016, § 516 BGB para. 24, in Säcker et al. (eds.), Münchener Kommentar zum BGB, Vol. III (7th ed.) Chiusi, T. J., 2013, § 516 BGB para. 49, in J. von Staudingers Kommentar zum BGB mit Einführungsgesetz und Nebengesetzen.

[30] BGH Oct. 2, 1991, 45 Neue Juristische Wochenschrift (NJW) 238, 239 (1992); Koch (2016) supra note 29 at § 516 para. 25; Chiusi (2013) supra note 29 at § 516 para. 40.

[31] For a similar assessment, see Bareiß (2012) supra note 20 at 461.

[32] Under German private law, sponsoring contracts are construed as a combination of service and work and labor contracts, see Schaub, R., 2008, Sponsoring und andere Verträge zur Förderung überindividueller Zwecke, Mohr Siebeck, 208-11.

[33] Bareiß (2012) supra note 20 at 461.

[34] Koch (2016) supra note 29 at § 516 para. 14, 24; Chiusi (2013) supra note 29 at § 516 para. 49.

[35] For this majority view see, for instance, BGH Nov. 27, 1991, 116 BGHZ 167 (170); Koch (2016) supra note 29 at § 516 para. 27. For an overview of the development of the doctrine, see Fischer, M., 2002, Die Unentgeltlichkeit im Zivilrecht, Heymann, 42-44; for a critique see Harke, J. D., 2018, § 516 BGB para. 67-70, in Gsell, B., W. Krüger, S. Lorenz, and C. Reymann (eds.), Beck'scher Online Grosskommentar zum BGB, Verlag C. H. Beck München.

[36] Bareiß (2012) supra note 20 at 461; Jansen and Pfeifle (2012) supra note 17 at 1843.

[37] If transactions are qualified as "mixed donations," see for instance Jansen and Pfeifle (2012) supra note 17 at 1843 note 6, these classifications are misguided. Only if the discrepancy in value was indeed reverse, i.e., the consideration was worth more than the worth more than the contribution, the transaction would be treated as a mixed donation, that is a single contract that combines elements of a donation and a sale (negotium mixtum con donatione), BGH Sep. 23, 1981, 82 BGHZ 274 (281-2); Koch (2016) supra note 29 at § 516 para. 34; Chiusi (2013) supra note 29 at § 516 para. 63. However, as a matter of pure economic rationality, it is almost inconceivable that the value of the consideration is higher than the crowdfunding contribution, because then the campaign would lose money and miss its primary purpose.

[38] Koch (2016) supra note 29 at § 516 para. 34; Harke (2018) supra note 35 at § 516 para. 104-5; Hähnchen, S., 2017, § 516 para. 28, in Westermann, H. P., B. Grunewald, G. Maier-Reimer; B. Erman, and W. Erman (eds.), Erman Bürgerliches Gesetzbuch, 15th ed., ottoschmidt; Dellios, G., 1981, Zur Präzisierung der Rechtsfindungsmethode bei "gemischten" Verträgen, Mohr Siebeck, 103-4 ; for a critique see Ernst, W., 2010, "Entgeltlichkeit - Eine Untersuchung am Beispiel des Tauschs, der gemischten Schenkung und anderer Verträge," in Richardi R., J. Wilhelm, and T. Lobinger (eds.), Festschrift für Eduard Picker, Mohr Siebeck, 139, 170-171.

[39] On the precise legal qualification of the latter, see infra 7.

[40] Bareiß (2012) supra note 20 at 461. If the campaign pertains to the production of a tangible good, the contract is one for work and materials ("Werklieferungsvertrag"), to which sales law also applies, cf. BGB, § 650 s. 1.

[41] For the general qualification of the acquisition of purchase rights for a consideration as a sale of rights, see Harm Westermann, P., 2016, § 453 BGB para. 4, in Säcker et al. (eds.), Münchener Kommentar zum BGB, Vol. III, 7th ed.

[42] id.

[43] See supra B.III.1.

[44] Renner (2014) supra note 7 at 263; see also Berger, S. C., and B. Skiera, "Elektronische Kreditmarktplätze: Funktionsweise, Gestaltung und Erkenntnisstand bei dieser Form des "Peer-to-Peer Lending"," 45 Kredit und Kapital 289, 291 (2012) (showing that a leading platform foresees that acquired cars are used as collateral in auto loans).

[45] See infra 0.

[46] Renner (2014) supra note 7 at 264.

[47] Sometimes, in order to minimize the need for costly information sharing, the credit-extending bank sells and assigns the redemption claim to a servicing firm, which is linked to the platform and subsequently passes the claim on to the ultimate investor, Renner (2014) supra note 7 at 264.

[48] In the U.S., the same result is reached with synthetic notes that replicate the cash flows from the loans platforms themselves and that are acquired by investors who may also trade on a secondary market, see Marte, J., 2010, "Credit crunch gives 'microlending' a boost," Wall Street Journal, September 26, at 1.

[49] For anecdotal evidence in addition to the empirical findings reported infra, see Jansen and Pfeifle (2012) supra note 17 at 1844; Bareiß (2012) supra note 20 at 461.

[50] Klöhn et al. (2016), supra note 12 at 145.

[51] The following section reiterates the main findings in Klöhn et al. (2016) supra note 12 at 148-178.

[52] Id. at 149 and 152-4 (showing that contracts usually are loan agreements that can be terminated after 5-7 years after a minimum notice period has elapsed and automatic termination after a fixed contract term representing an exception).

[53] Id. at 155-6 [finding annual interest rates varying from 1% to 8% and due either upon redemption, or periodically (annually, quarter-annually)].

[54] Id. at 158-60 (identifying an unlimited pro-rata profit participation in four-fifths of the cases and a capped participation in others).

[55] Id. at 161-5 (describing that investors either receive a payment based on an appraisal of the issuer at the time the investment is terminated or a fraction of the proceeds that accrue to equity holders if they sell their shares)

[56] Id. at 160 (also showing that liability was sometimes not limited in the past).

[57] Id. at 177-8. The reason for the subordination comes from prudential banking regulation, which would submit borrowers to an authorization requirement if the loans were not subject to a specific subordination clause, see Pölzig, D., 2014, "Nachrangdarlehen als Kapitalanlage," 68 Wertpapier-Mitteilungen (WM) 917, 919.

[58] Id. at 166-8 (indicating that contracts provide for a proportional adjustment of the participation ratio under which losses can only occur if the issuer is undervalued in the new round of financing).

[59] Id. at 168.

[60] Id. at 168-73 (describing that disclosure obligations provide inter alia for quarterly reporting, disclosure of annual accounts, and overview of profit and revenue participation).

[61] Id. at 173-76; Jansen and Pfeifle (2012) supra note 17 at 1844.

[62] For instance, Jansen and Pfeifle (2012) supra note 17 at 1846; Bareiß (2012) supra note 20 at 461; Klöhn et al. (2016) supra note 12 at 145; on the development of the market, which clearly shifted towards subordinated profit participating loans in reaction to prudential regulation see Klöhn et al. (2016) supra note 15 at 58-9.

[63] See generally, Handelsgesetzbuch [HGB] [Commercial Code], May 10, 1897, RGBl. 219, §§ 230-6 HGB, translation at https://bit.ly/2QmUS7K.

[64] The latter have not received a special treatment neither in the German Civil nor the Commercial Code, but are anticipated in different legislative acts, like for instance Capital Investment Act [Vermögensanlagengesetz, VermAnlG], Dec. 6, 2011, BGBl. I at 2481, § 1 para. 2 nr. 4, https://bit.ly/2QphGUl or AktG, § 221 paras. 3 and 4. The lack of statutory prescriptions together with the fundamental principle of freedom of contract allow for a highly flexible individual design of parties' obligations in these profit participation rights.

[65] The contract combines a regular loan with an additional stipulation of sharing in the profits or sales that flow from the investment of the borrowed funds as compensation for the lender, see for instance Schmidt, K., 2012, § 230 HGB para. 54, in Schmidt, K., (ed.), Münchener Kommentar zum HGB, Verlag: Beck, vol. 3, 3rd ed.; Schäfer, C., 2017, Vor § 705 BGB para. 107, in Säcker et al. (eds.), Münchener Kommentar zum BGB, Verlag: Beck, vol. 6, 7th ed.; Huffer, H., 1970, Das partiarische Geschäft als Rechtstypus.

[66] RG May 11, 1920, 99 RGZ 161 (163), BGH Jun. 6, 1965, 19 WM 1965, 1052 (1053); Freitag, R., 2015, § 488 BGB para. 70, in J. von Staudingers Kommentar zum BGB mit Einführungsgesetz und Nebengesetzen; Blaurock, U., 2010, Handbuch der Stillen Gesellschaft , 7th ed., para 8.30.

[67] See HGB § 233 para. 1, which describes typical control rights of a silent partner.

[68] See generally, Habersack, M., 2016, § 221 AktG para. 93, in Goette, W., and M. Habersack (eds.), Münchener Kommentar zum AktG, C. H. Beck, vol. 4, 4th ed.; Merkt, H., 2016, § 221 AktG para. 46, in Schmidt, K., and M. Lutter (eds.), AktG, vol. 2, 3d ed..

[69] Jansen and Pfeifle (2012) supra note 17 at 1846; Bareiß (2012) supra note 20 at 461.

[70] See generally, Schmidt, K., 2012, § 230 HGB para. 54, in Schmidt, K., (ed.), Münchener Kommentar zum HGB, vol. 3, 3rd ed.; Schäfer, C., 2017, Vor § 705 BGB para. 107, in Säcker et al. (eds.), Münchener Kommentar zum BGB, vol. 6, 7th ed.; in the context of crowdinvesting Jansen and Pfeifle (2012) supra note 17 at 1846; Bareiß (2012) supra note 20 at 461.

[71] See Jansen and Pfeifle (2012) supra note 17 at 1846 (submitting that these transactions should be seen as profit participation contracts sui generis).

[72] Cf. BGB § 488 para. 1 sentence 2 declaring a repayment obligation essential part of a loan contract under German law, see also Freitag (2015) supra note 66 at § 488 para. 70. Where investors receive a fraction of each individual sale (for instance a share of each cinema ticket sold), this contract design can be qualified as a special annuity arrangement (until the contribution is repaid) and a profit participation (after repayment).

[73] Bareiß (2012) supra note 20 at 462. Given the legal character of the participation rights as debt contracts, such a classification is not excluded simply because investors do not receive an ownership stake in the funded entity (for this view see Jansen and Pfeifle (2012) supra note 17 at 1844-5).

[74] On the lack of legal innovation in pertinent respect, see already supra A.I.

[75] See supra B.III.4.

[76] See infra A.I.1.b).

[77] Regulation (EC) No 593/2008 of the European Parliament and of the Council of 17 June 2008 on the law applicable to contractual obligations (Rome I), 2008 O.J. (L 177) 6 [hereinafter Rome I Regulation].

[78] According to the German majority view this requires securitization of the claim to facilitate its transfer by an assignment of the instrument, see for instance Martiny, D., 2012, Art. 1 Rom I VO para. 58-9, in Säcker et al. (eds.), Münchener Kommentar zum BGB, vol. 12, 7th ed.; Kieninger, E. M., 2018, VO (EG) 593/2008 Art. 1 para 15, in Ferrari et al. (eds.), Internationales Vertragsrecht, 3rd ed.

[79] Martiny (2012) supra note 78 at Art. 1 Rom I VO paras. 65, 72; Kieninger (2018) supra note 78 at VO (EG) 593/2008 Art. 1 para 15; Wedemann, F., § 230 para. 119, in: Oetker, H., (ed.) HGB, 5th ed.; Roth, W-H., 2014, Internationalprivatrechtliche Aspekte der Personengesellschaften, 43 Zeitschrift für Unternehmens- und Gesellschaftsrecht (ZGR) 168, 179; Spindler, G., 2017, "Crowdfunding und Crowdinvesting – Sach- und kollisionsrechtliche Einordnung sowie Überlagerung durch die E-Commerce-Richtlinie," 29 ZBB 129, 139.

[80] Rome I Regulation, art. 1 para. 2 lit. f).

[81] Rome I Regulation, art. 3 para. 1. On the preconditions for valid clauses in standard terms, see Verein für Konsumenteninformation v. Amazon, Case C-191/15, [2016] ECLI:EU:C:2016:612 (delivered July 28, 2016).

[82] The anecdotal evidence reported in Spindler (2017) supra note 79 at 139 note 144 is inconclusive as the terms and conditions the author cites are those for the relationship between investors and platforms only.

[83] Rome I Regulation, art. 6 para. 2 s. 2.

[84] According to recital 30 of the Rome I Regulation, its art. 6 para. 4 lit. d) pertains only to financial instruments within the meaning of Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU, art. 4 para. 17 and Annex I C, [2014] O.J. (L 173) at 349, which does neither capture silent partnership interests, nor profit participation rights, nor subordinated profit participating loans, for a discussion see Spindler (2017) supra note 79 at 139.

[85] Infra A.I.1.b). See also Spindler (2017) supra note 79 at 141 (showing that the relevant rules cannot be qualified as overriding mandatory provisions within the meaning of Rome I Regulation, art. 9 para. 1).

[86] Rome I Regulation, art. 6 para. 1 points to the law of the country of the consumer's habitual residence.

[87] Rome I Regulation, art. 3 para. 1.

[88] For a specific discussion, albeit focused on crowdinvesting relationships, see Spindler (2017) supra note 79 at 140.

[89] Otherwise this intermediate step in the discharge of the platform's role can be seen as "deposit business," which requires a banking license, cf. infra D.I.1.

[90] See Banking Act [Kreditwesengesetz, KWG], Sep. 9, 1998, BGBl. I at 2446, § 32, with the relevant definitions codified in KWG, § 1 para. 1a sentence 2 and Securities Trading Act [Wertpapierhandelsgesetz, WpHG], Sep. 9, 1998, BGBl. I at 2708, § 2 para. 3, https://bit.ly/2luhReh.

[91] BaFin, 2007, Merkblatt zur Erlaubnispflicht der Betreiber und Nutzer einer internetbasierten Kreditvermittlungsplattform nach dem KWG (2007), https://bit.ly/2NgoB0j; Renner (2014) supra note 7 at 264.

[92] As defined in KWG § 1 para. 1a sentence 2 no. 1; for an extensive discussion of the latter aspect see Veith, J., 2016, "Crowdlending – Anforderungen an die rechtskonforme Umsetzung der darlehensweisen Schwarmfinanzierung," 16 Zeitschrift für Bank- und Kapitalmarktrecht (BKR) 184, 186-7.

[93] Defined in KWG, § 1 para. 2 No. 1.

[94] Schäfer, F. A., 2016, § 1 KWG para. 46, in Boos, K-H., R. Fischer, and H. Schulte-Mattler (eds.), KWG, CRR-VO, 5th ed.

[95] Renner (2014) supra note 7 at 265. For a general discussion of the respective element in KWG, § 1 para. 2 No. 1 see Demgensky, S., and A. Erm, 2001, "Der Begriff der Einlage nach der 6. KWG-Novelle," 55 WM 1445, 1448.

[96] As defined in KWG, § 1 para. 2 No. 2.

[97] For the U.S. model which sees platforms extend loans see supra note 48.

[98] KWG, § 37 para. 1 s. 4 empowers BaFin to stop the operations of and wind-down firms that were involved in the initiation, conclusion, or execution of prohibited (unauthorized) activities.

[99] Veith (2016) supra note 92 at 186; Renner (2014) supra note 7 at 266 (with a critique that favors a more restrictive interpretation); see also Schwennicke, A., 2010, "Vergabe privater Darlehen und Erlaubnispflicht nach dem KWG," WM 542, 548 (2010) arguing in favor of firm de minimis limits.

[100] Cf. KWG, § 32 para. 1 s. 1. As a matter of practice, the second alternative of the provision, that the activity requires a commercial business organization (in kaufmännischer Weise eingerichteter Geschäftsbetrieb) is mute, because the elements of a commercial activity are usually met, even though no specific organizational arrangements are necessary.

[101] BGH Jul. 11, 2006, 60 Der Betrieb (DB) 2061, 2062 (2006); BaFin, Merkblatt Kreditgeschäft (2016), https://bit.ly/2Rc0Ppn; Schäfer (2016) supra note 94 at § 1 KWG para. 22.

[102] BaFin supra note 101.

[103] For an alternative proposal that would retain a direct contracting model but use subordinated loans see Veith (2016) supra note 92 at 187.

[104] BaFin supra note 103; Schäfer supra note 94 at § 1 KWG para. 43; Heer, P. E., 2012, "Die Übertragung von Darlehensforderungen – eine systematische Übersicht – zugleich Anmerkung zum Urteil des BGH vom 19. 4. 2011 – XI ZR 256/10," 12 BKR 45, 47 (2012).

[105] For an extensive discussion see Veith (2016) supra note 92 at 188-9.

[106] For an overview on the question if issuers need an authorization because they engage in "deposit business" (Einlagengeschäft) within the meaning of KWG, § 1 para. 2 No. 1, see for instance Nietsch, M., and N. Eberle, 2014, "Bankaufsichts- und prospektrechtliche Fragen typischer Crowdfunding-Modelle," 67 Der Betrieb (DB) 1788, 1790.

[107] Supra D.I.1.

[108] WpHG § 2 para. 2b; KWG § 1 para. 11

[109] Capital Investment Act [Vermögensanlagengesetz, VermAnlG], Dec. 6, 2011, BGBl. I at 2481, § 1 para. 2, https://bit.ly/2QphGUI.

[110] Prior to the 2015 reforms a debate existed about whether the definition of financial assets also included profit participating loans. See, for instance, Weitnauer, W., and J. Parzinger, 2013, "Das Crowdinvesting als neue Form der Unternehmensfinanzierung," 4 Gesellschafts- und Wirtschaftsrecht (GWR) 153, 155 (advocating an inclusive definition on normative grounds); Nietsch and Eberle (2014) supra note 106 at 1790 and 1793 (opposing such a wide definition).

[111] As defined in KWG, § 1 para. 1 sentence 2 no. 10; WpHG, § 2 para. 3 sentence 1 no. 5.

[112]On the general precondition of a firm underwriting to fall under the statutory regime see Schäfer (2016) supra note 94 at § 1 KWG para. 112; Kumpan, C., 2010, § 2 WpHG para. 72, in Schwark, E., and D. Zimmer (eds.), Kapitalmarktrechtskommentar, 4th ed..

[113]As defined in KWG § 1 para. 1a sentence 2 no. 1c; WpHG, § 2 para. 3 sentence 1 no. 6.

[114]BaFin, 2009, "Merkblatt – Hinweise zum Tatbestand des Platzierungsgeschäfts," https://bit.ly/2NgoB0j, (requiring a disclosed open agency relationship).

[115]Klöhn and Hornuf (2015) supra note 4 at 249-50.

[116]For a detailed description of the relevant provisions' content see, Jansen and Pfeifle (2012) supra note 17 at 1850-1; Klöhn and Hornuf (2015) supra note 4 at 250-1.

[117]As defined in KWG § 1 para. 1a sentence 2 no. 1; WpHG, § 2 para. 3 sentence 1 no. 4.

[118]See, for instance, Nietsch, M., and N. Eberle, 2014, "Crowdinvesting – Welche Auswirkungen hat das geplante Kleinanlegerschutzgesetz?" 67 DB 2575, 2576.

[119]BaFin, 2011, "Merkblatt – Hinweise zum Tatbestand der Anlagevermittlung,"; Assmann, H-D., 2012, § 2 WpHG para. 81, in Assmann, H-D., and U. H. Schneider (eds.), WpHG, 6th ed.

[120]As defined in KWG, § 1 para. 6 no. 8 Buchst. e); WpHG, § 2a para. 1 no. 7 Buchst. e).

[121]Trade Regulation [Gewerbeordnung, GewO], Feb. 22, 1999, BGBl. I at 202, § 34 para. 1 sentence 1, https://bit.ly/2y4Opa7.

[122]As defined in KWG, § 1 para. 1a no. 1b; WpHG, § 2 para. 3 no. 8. See also Directive 2004/39/EC of the European Parliament and of the Council of 21 April 2004 on markets in financial instruments amending Council Directives 85/611/EEC and 93/6/EEC and Directive 2000/12/EC of the European Parliament and of the Council and repealing Council Directive 93/22/EEC, 4 para. 1 no. 15, 2004 O.J. (L 145) 1 [hereinafter MiFiD].

[123]Klöhn & Hornuf, supra note 4 at 251.

[124]See MiFiD, recital 5.

[125]BaFin (2011) supra note 119; BaFin, Crowdinvesting, https://bit.ly/2QpcZtl

[126]For a general overview see Assmann (2012) supra note 119 at para. 110.

[127]Regulation (EU) No 575/2013 of the European Parliament and of the Council of 26 June 2013 on prudential requirements for credit institutions and investment firms and amending Regulation (EU) No 648/2012, [2013] O.J. (L 176) 1.

[128]WpHG, §§ 63-98.

[129]Council Regulation (EU) No 1024/2013 of 15 October 2013 conferring specific tasks on the European Central Bank concerning policies relating to the prudential supervision of credit institutions, art. 4 para. 1 lit. a), [2013] O.J. (L 287) 63.

[130]BGB, § 491a. For detailed description of the information duties Schürnbrandt, J., 2016, Vor § 491 para. 4, in Säcker et al. (eds.), Münchener Kommentar zum BGB, vol. 3, 7th ed.; for a discussion in the context of crowdlending see Veith (2016) supra note 92 at 193; Renner (2014) supra note 7 at 268-9.

[131]Directive 2008/48/EC of the European Parliament and of the Council of 23 April 2008 on credit agreements for consumers and repealing Council Directive 87/102/EEC, art. 5-7, [2008] O.J. (L 133) 66, implemented in BGB, §§ 491a, 493.

[132]Einführungsgesetz zum Bürgerlichen Gesetzbuche (EGBGB) [Introductory Act to Civil Code], Aug. 18, 1896, RBl. 604, Art. 247 § 13.

[133]Gilson, R. J., 2003, "Engineering a venture capital market: lessons from the American experience," 55 Stanford Law Review 1067, 1077 (2003) (describing the key problems in venture capital investing); Cumming, D. J., and S. A. Johan, 2009, Venture capital and private equity contracting: an international perspective, Elsevier, 48-52 (showing that the features of equity claims make for lemon markets in both equity and debt financing of start-up firms because unprofitable ventures are more likely to issue equity while riskier ones have a proclivity to seek debt financing).

[134]Bradford, C. S., 2012, "Crowdfunding and the Federal Securities Laws," 2012 Columbia Business Law Review, 1, 105; more specifically for Germany Meschkowski and Wilhelmi (2013) supra note 4 at 1410-11.

[135]An investment advisor has to provide recommendations inter alia with a view to the specific financial situation of individual investors and the characteristics of the investment, see for instance BGH, July 6, 1993, BGHZ 123, 126 (128-9); Emmerich, V., 2016, § 311 BGB para. 101, in Säcker et al. (eds.), Münchener Kommentar zum BGB, vol. 2, 7th ed.

[136]However, if platforms use client data to provide recommendations derived from algorithms, for instance based on past investment behavior, they might be seen as investment advisors and incur far reaching fiduciary obligations, see Jansen and Pfeifle (2012) supra note 17 at 1849.

[137]Jansen and Pfeifle (2012) supra note 17 at 1849; generally on the highly relevant distinction of investment advice on the one hand and investment brokerage on the other in German law, see Assmann, H-D., 2002, "Negativberichterstattung als Gegenstand der Nachforschungs- und Hinweispflichten von Anlageberatern und Anlagevermittlern," 23 ZIP 637, 648; Möllers, T. M. J., and T. Ganten, 1998, "Die Wohlverhaltensrichtlinie des BAW im Lichte der neuen Fassung des WpHG – Eine kritische Bestandsaufnahme," 27 ZGR 773, 785-6.

[138]For these well-established, general preconditions for a tacit agreement to provide information, see for instance BGH, Mar. 22, 1979, 74 BGHZ 103 (106); BGH, Mar. 4, 1987, 100 BGHZ 117 (118-9); BGH, May 13, 1993, 14 ZIP 997 (1993) BGH, Oct. 19, 2006, 27 ZIP 2221 (2006); Heermann, P. W., 2017, § 675 BGB para. 122, in Säcker et al. (eds.), Münchener Kommentar zum BGB, vol. 5/2, 7th ed.; Siol, J., 2017, § 45, "Anlagevermittlung und Prospekthaftung der Banken," in Schimansky et al. (eds.), Bankrechts-Handbuch para. 6, 5th ed.

[139]Meschkowski and Wilhelmi (2013) supra note 4 at 1413.

[140]For an analysis and critique of cases, see Benedict, J., 2005, Die Haftung des Anlagevermittlers, 26 ZIP 2129, 2131-3.

[141]See generally 74 BGHZ 103 (110); BGH, Feb. 16, 1981, 80 BGHZ 80 (81-2); Emmerich (2016) supra note 135 at para 127; Siol (2017) supra note 138 at para 9.

[142]Emmerich (2016) supra note 135 at para 127.

[143]Emmerich (2016) supra note 135 at para 127; Siol (2017) supra note 138 at para 9.

[144]Jansen and Pfeifle (2012) supra note 17 at 1850.

[145]VermAnlG, § 18 para. 1 no. 2 and no. 3 empower the supervisor (BaFin) to prohibit public offerings of investments from going forward if they violate the prospectus requirements.

[146]See Meschkowski and Wilhelmi (2013) supra note 4 at 1415; Klöhn and Hornuf (2012) supra note 4 at 259; Nietsch and Eberle (2014), supra note 106 at 2579.

[147]Small Investor Protection Act [Kleinanlegerschutzgesetz, KASG], July 3, 2015, BGBl. I at 1114, art. 2 no. 4.

[148]VermAnlG, § 2a.

[149]For a more granular description of the relevant statutory requirements, Klöhn et al. (2016) supra note 15 at 59-60; for in-depths analyses see Casper, M., 2015, "Das Kleinanlegerschutzgesetz – zwischen berechtigtem und übertriebenem Paternalismus," 27 ZBB 265, 275-80; Nietsch and Eberle (2014), supra note 106 at 1789.

[150]For a policy discussion of investment limits, see Klöhn & Hornuf (2012) supra note 4 at 262-4. For a critique of the current limits see Klöhn, L., and L. Hornuf, 2015, "Die Regelung des Crowdfunding im RegE des Kleinanlegerschutzgesetzes – Inhalt, Auswirkungen, Kritik, Änderungsvorschläge," 68 DB 47, 52-53.

[151]VermAnlG, §§ 13, 15. For details see Klöhn et al. (2016) supra note 15 at 60.

[152]VermAnlG, § 13 para. 3a.

[153]VermAnlG, § 16 para. 1; see Klöhn et al. (2016) supra note 15 at 60.

[154]VermAnlG, § 12 para. 2 and 3 prescribe that the expressed warnings that a total loss of funds invested is possible and that a promised return is not guaranteed are sufficiently visible also in advertisement campaigns. For a granular delineation of the restrictions see Waschbusch, G., 2016, "Die Masse macht's – Crowdfunding als Finanzierungsmöglichkeit für Existenzgründer," 67 Der Steuerberater (StB) 206, 208.

[155]On the concept, see Romano, R., 1998, "Empowering investors: a market approach to securities regulation," 107 Yale Law Journal 2359, 2362, 2418 (1998) (proposing that issuers be permitted to opt into both U.S. States' and foreign nations disclosure regimes); Palmiter, A. R., 1999, "Toward disclosure choice in securities offerings," Columbia Business Law Review 1, 86-91 (restricting issuer choice to the selection of a primary market disclosure regime); for a critique see Fox, M. B., 1999, "Retaining mandatory securities disclosure: why issuer choice is not investor empowerment," 85 Virginia Law Review 1335, 1345-56 (holding that the divergence between managers' private benefits and social benefits derived from disclosure rules will induce suboptimal outcomes under a regime of issuer choice).

[156]Jansen and Pfeifle (2012) supra note 17 at 1844; Bareiß (2012) supra note 20 at 459 (reporting that payment accounts of contributors are only debited if target levels for overall financing are reached or contributions are returned if these levels are undercut).

[157] At least the general duty to avoid any acts that threaten the purpose parties pursue with the contract (on the respective construction of the accompanying duties mentioned in BGB, § 241 para. 2 see Bachmann, G., 2016, § 241 para. 85, in Säcker et al. (eds.), Münchener Kommentar zum BGB, vol. 2, 7th ed.) applies, regardless of the legal qualification of the crowdfunding relationship. See Jansen and Pfeifle (2012) supra note 17 at 1845 note 21, 1846 without doctrinal specification. See also infra E.III.

[158] BGB, §§ 280 para. 1, 276 para. 1.

[159] The cases do not specifically pertain to crowdinvesting, but to investment brokers in general and are, therefore, relevant for the determination of platforms' duties to inform. For an overview see Siol (2017) supra note 138 at para. 18-21.

[160] See already supra note 157.

[161] To find negligence, BGB, § 276 para. 2, requires a showing that the debtor violated the duty of care as observed by the respective public circles. Hence, the objective standard needs to be specified with a view to the respective contractual obligation, see for instance BGH, Mar. 17, 1981, BGHZ 80, 186 (193); Grundmann, S., 2016, § 276 BGB para. 55-6, in Säcker et al. (eds.), Münchener Kommentar zum BGB, vol. 2, 7th ed.).

[162] An explicit safe harbor protecting business judgement against judicial second guessing prone to hindsight bias is codified in AktG, § 93 para. 1 s. 2 for managers of stock corporations. Beyond the narrow scope of this specific provision, the underlying principle is also relevant in general private law.

CAPCO

THE FUTURE. NOW.

ABOUT CAPCO

Capco is a global technology and management consultancy dedicated to the financial services industry. Our professionals combine innovative thinking with unrivalled industry knowledge to offer our clients consulting expertise, complex technology and package integration, transformation delivery, and managed services, to move their organizations forward. Through our collaborative and efficient approach, we help our clients successfully innovate, increase revenue, manage risk and regulatory change, reduce costs, and enhance controls. We specialize primarily in banking, capital markets, wealth and investment management, and finance, risk & compliance. We also have an energy consulting practice. We serve our clients from offices in leading financial centers across the Americas, Europe, and Asia Pacific.

To learn more, visit our web site at **www.capco.com, or follow us on Twitter, Facebook, YouTube and LinkedIn.**

WORLDWIDE OFFICES

APAC
Bangalore
Bangkok
Hong Kong
Kuala Lumpur
Pune
Singapore

EUROPE
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
London
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA
Charlotte
Chicago
Dallas
Houston
New York
Orlando
Toronto
Tysons Corner
Washington, DC

SOUTH AMERICA
São Paulo

WWW.CAPCO.COM