Henley Business School
UNIVERSITY OF READING

CAPCO

JOURNAL

THE CAPCO INSTITUTE JOURNAL OF FINANCIAL TRANSFORMATION

AUTOMATION

Henley Business School – Capco Institute
Paper Series in Financial Services

#46
11.2017

THE HENLEY MBA

Realise your potential with an MBA
from a triple-accredited business school.

hly.ac/the-henley-mba

Henley
Business School
UNIVERSITY OF READING

Where business comes to life

AACSB ACCREDITED

EFMD EQUIS ACCREDITED

ASSOCIATION OF MBAS ACCREDITED

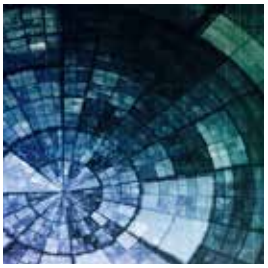
JOURNAL

THE CAPCO INSTITUTE JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

CONTENTS

AUTOMATION



BUSINESS MODELS

INVESTMENTS

DEAR READER,

Welcome.

Intelligent automation is one of the most prominent trends in financial services, fundamentally redefining the industry landscape as we know it.

As the financial services industry accelerates towards a digital universe, advances in data, artificial intelligence and robotic process automation offer much more than purely cost savings.

Robotic process automation – RPA – is a set of technologies which enables the automation of routine activities. These robotic, rules-based solutions are well-suited for banks responding to increasing regulatory and compliance demands; they offer more robust responses and liberate the human workforce to concentrate on truly value-adding activities.

This can offer attractive benefits for financial institutions. However, choosing automation technology should not be automatic. Instead, to realize the full potential of RPA, institutions must form a very clear understanding of how to optimize the benefits inherent in this solution and the business case, and then crucially, how to manage these solutions once implemented.

In this edition of the Journal, published in association with the world-class Henley Business School, we have included papers that focus on the origins and status of RPA. We identify the steps needed to avoid early stage implementation pitfalls, and offer insights into transforming the technical potential of RPA into real, sustained advantage.

This edition also considers several industry challenges. In our Business Models and Investments sections, you will find perspectives covering topics from risk management transformation to financial planning.

Once again, the breadth of topics included in this edition confirms the complexity of our industry. RPA is a powerful transformation tool, but a complicated one, and the key to success is to take proper steps to correctly evaluate and execute the appropriate implementation strategy.

I hope the articles provided in this edition of the Journal continue to inform your strategic technology choices, and I wish you every success in an increasingly automated future.

LANCE LEVY
CEO, Capco

AUTOMATION

PREVIOUS EDITIONS OF THE CAPCO JOURNAL OF FINANCIAL TRANSFORMATION ARE AVAILABLE AT **WWW.CAPCO.COM/INSTITUTE**

Regtech as a new legal challenge

Bridging the gap between investment banking infrastructure and distributed ledgers

Rethinking robotics? Take a step back

To robo or not to robo: The rise of automated financial advice

Understanding robotic process automation (RPA)

Robotizing Global Financial Shared Services at Royal DSM

The financial auditing of distributed ledgers, blockchain, and cryptocurrencies

Targeting the robo-advice customer: The development of a psychographic segmentation model for financial advice robots

Regtech as a new legal challenge

ROLF H. WEBER | Professor for Civil, Commercial and European Law,
University of Zurich Law School, and Counsel, Bratschi Wiederkehr & Buob AG (Zurich)

ABSTRACT

In view of the digitization of many parts of the economy, regtech will gain importance and constitute a growing challenge for private organizations as well as for regulators. As far as businesses are concerned, the new technological environment requires the implementation of better risk management and IT compliance systems. Regulators should provide an example of a forward-looking implementation of IT-enabled systems.

A paradigm shift is expected to occur in financial regulation. Among others, the well-known principle of "know-your-customer" (KYC) might be complemented by a new principle, "know-your-data" (KYD). This new perspective leads to more data-driven regulations and compliance models.

Since the challenges rapidly evolve in the technological environment, regtech should be built into a reconceptualized (financial) regulatory regime. The paradigm shift based on a holistic approach will have to put more emphasis on the nexus of data and digital identity. Data needs to be gathered globally on a real-time basis and processed in a coherent way in order to meet the cross-border needs of businesses and regulators.

1. REGTECH: TERM AND MEANING

Officially, the term "regtech" was used for the first time in March 2015 in the U.K., mainly related to the potential of fintech, by the U.K. Government Office for Science[1] and by the British Finance Ministry.[2] Since then, regtech has become a buzzword for private enterprises as well as public regulatory authorities. Nevertheless, there is not a universally accepted definition of regtech; instead we have a number of different descriptions.[3]

In the U.K., regtech is usually linked to fintech; it is even argued that fintech encompasses regtech.[4] Such an approach appears to be problematic: regtech has the potential for application in a wide range of contexts; for example, the potential of continuous monitoring capacity, of providing close to real-time insights, of using deep learning and artificial intelligence, and of enabling the functioning of the markets nationally and globally.[5] Even if regtech might remain a major driver for fintech, the two terms need to be distinguished since regtech represents more than just an efficiency tool; rather it is a pivotal change leading to a paradigm shift in regulation [Arner et al. (2017), 382 and 383].

By harnessing the capabilities of new technologies such as big data, cloud computing, and distributed ledger technologies, regtech is able to design solutions to help enterprises and supervisory authorities across all sectors of the economy to ensure that they comply with the given regulatory environment.[6] In this sense, the contours of regtech can be described as follows:[7]

- **Regtech in a narrow sense:** organizational compliance processes of (financial) enterprises can be designed and implemented in a more efficient way by using new technologies through, for example, the replacement of analog by digital processes. Automation is suitable for facilitating the reporting requirements and the supervision of regulated enterprises.
- **Regtech in a wide sense:** if regulatory and supervisory authorities also apply digital technologies for the execution of their activities, better efficiency objectives are achievable; thereby, coherence between public and private technologies becomes essential. In other words, unified data formats, compatible programming interfaces (API), and machine readable interactions are needed.[8]

Regtech can be understood as a new regulatory program and could deal with everything from digital identity to data sovereignty, thereby extending beyond the financial sphere.

2. REGTECH: IMPORTANCE AND SCOPE OF APPLICATION

From a business perspective, regtech is an answer to the increasing administrative tasks caused by the growing number of regulations that have been released during the last ten years since the outbreak of the financial crisis.

2.1. Economic relevance

According to a survey, regulators and authorities of the G-20 have implemented about 60,000 regulatory ordinances.[9] The costs following the observance of and compliance with the new regulations might exceed U.S.$ 100 billion in the financial services sector.[10] Approximately 10% to 15% of all working places of financial intermediaries are devoted to functions of governance, risk management, and compliance.[11] Overall, banks spend 3% to 5% of their total costs on compliance efforts [IIF (2015), 1].

The increasing complexity of the regulations and the subsequent operational risks in case of non-compliance requires that financial intermediaries invest high amounts for compliance matters. The fines paid by the banks between 2010 and 2014 have grown by the factor 45;[12] hence, avoidance of such consequences might be cheaper than the increased compliance costs. Looking from this perspective, the bank crisis has caused a compliance crisis.

The reason why regtech has become an important issue can be attributed to (i) the post-crisis regulation

[1] U.K. Government Office for Science (Sir Mark Walport, Government Chief Scientific Adviser), 2015, "FinTech futures: the U.K. as a world leader in financial technologies," March 13, (http://bit.ly/2yihTDV), 5 et seq.: "FinTech has the potential to be applied to regulation and compliance to make financial regulation and reporting more transparent, efficient and effective – creating new mechanisms for regulatory technology, Regtech".
[2] HM Treasury's 2015 Budget Report, March 18, 53 (Section 1.204), 98 (Section 2.272).
[3] For an overview of publications related to "The birth of Regtech" see Robinson, M., 2016, "The Regtech ecosystem: in depth analysis (part 3 of 3)", March 10, http://bit.ly/2xROAqr.
[4] See for example Financial Conduct Authority (FCA), 2015, "Call for input: supporting the development and adoption of Regtech," November, http://bit.ly/2fE7zLb, 3; a detailed overview is given by Arner, D. W., J. Barberis, and R. P. Buckely, 2017, "FinTech, Regtech, and the reconceptualization of financial regulation," Northwestern Journal of International Law & Business 37/3, 371, 381 et seq.
[5] Regtech has the potential for application in a wide range of contexts [Arner et al. (2017), 383].
[6] Larsen, K. S., and S. Gilani, 2017, "Regtech is the new black – the growth of Regtech demand and investment," Journal of Financial Transformation, 45, 22 et seq.
[7] Contratto, F., 2016, "Regtech": Digitale Wende für Aufsicht und Compliance, Weblaw-Jusletter 15 August, no. 6.
[8] Le Brocq, N., 2016, "Regtech – the new paradigm, recognizing the potential for technology to manage regulatory data and improve internal control and compliance," April 13, http://bit.ly/2x9fC8R.
[9] Simpson, D., 2017, "The collaborative path through the Regtech and FinTech jungles," April, 25, http://bitly/2pge7p1; for a detailed overview see Larsen and Gilani (2017), 23-25.
[10] Institute of International Finance (IIF), 2015, "Regtech: exploring solutions for regulatory challenges," Washington D.C., October 29, http://bit.ly/2yDgvrl, 1.
[11] BBVA Research, 2016, "Banking outlook," March 6, http://bit.ly/2fDaHGW, 14/15.
[12] Kaminski, P., and K. Robu, 2016, "A best-practice model for bank compliance," McKinsey & Co. http://bitly/2drDAVB; see also Larsen and Gilani (2017), 25.

requiring massive additional data disclosure from supervised entities, (ii) the developments in data science allowing the structuring of unstructured data (through artificial intelligence and deep learning), (iii) the economic incentives to minimize rising compliance costs, and (iv) the efforts of regulators to enhance the efficiency of supervisory tools [Arner et al. (2017), 383].

2.2 Scope of application

Regtech has many application fields. The main areas, due to the complex, fragmented, and ever-evolving financial regulatory regimes are:

- The most important area of regtech relates to **reporting** and **recordkeeping**. The availability of regtech products could make regulatory reporting easier and cheaper for market participants, and allow them to meet changing regulatory requirements [Larsen and Gilani (2017), 26/27]. Equally, market entrants might have fewer obstacles to offer their services if well-established regtech products and services are available. The newest technologies, in particular the distributed ledger approaches (DLT, for example blockchain),[13] may contribute to cost savings for companies. Instead of the traditional delivery of documents to the regulator for its review, DLT is in a position to provide regulators with almost instant access to transaction information. Experience shows that the use of new technologies can lead to significant cost savings by simplifying reporting and recordkeeping processes [Larsen and Gilani (2017), 27].
- Another area for regtech concerns **monitoring** and **surveillance**. Thereby, a starting point lies in the automated processing of data [Contratto (2016), no. 7/8]. The objective of these regulatory activities is to be seen in the identification and reduction of market abuse risks, apart from the notification of suspicious transactions to the relevant regulator [Larsen and Gilani (2017), 27].
- Often, the details of the monitoring activities, as well as the notification requirements, are not clearly stated in the applicable legal framework. Regtech could help to standardize the respective obligations and to provide means for cooperation in a responsive manner.[14]
- As far as banks are concerned, the **capital adequacy** and the **liquidity** requirements according to Basel III, CRD IV, Dodd-Frank, and Solvency II require a far-reaching data aggregation of risk information. Regtech appears to be an instrument that could support model simulations and scenario analyses that are typically applicable in the context of stress-tests [Arner et al. (2017), 393 and 395/96].
- An increasingly important number of regulatory provisions in the financial markets require the authentication and/or the identification of persons (customers) in view of compliance with the **KYC** requirements [Larsen and Gilani (2017, 25/26 and 27); Arner et al. (2017), 391/92 and 395]. Hence, the traditional paper-trail inquiries are increasingly replaced by the use of biometric identification methods (iris identification, scanning of finger prints). Regtech products could become a tool in the harmonization and/or standardization of the respective identification processes [Contratto (2016), no. 18].
- The number of automated applications in the context of **trade surveillance** [for example, with the purpose of securing best execution or to comply with risk parameters in the securities trade (margins)] gain importance in view of MiFID II, EMIR, or SEC Rule 15c3-5 [Arner et al. (2017), 408 and 410, Contratto (2016), no. 17].
- **Cybersecurity** issues have become another important topic in financial regulation. The vulnerability of existing frameworks requires the implementation of security measures [Arner et al. (2017), 400/01, Larsen and Gilani (2017), 27/28]. As outlined below, IT architecture issues will have to play a key role.

[13] Weber, R. H., 2017, "Regulatory environment of the ledger technology," Computer Law Review International 18:1, 1-6

[14] An obvious condition would be that international standard-setters are becoming more active.

3. REGULATORY ACTIVITIES

As mentioned, regtech has become a buzzword and caused international organizations and regulators to tackle the new developments.

3.1. International standardization organizations

The Financial Stability Board (FSB) and, to some extent, the International Monetary Fund (IMF), looked into regtech matters; the most important document addresses financial stability issues.[15] The Financial Action Task Force (FATF) and the Bank for International Settlements (BIS) have also been quite active;[16] IOSCO has equally issued a detailed report on regtech issues.[17] However, clear recommendations are not yet available.

A valuable contribution of the International Institute of Finance (IIF), a private organization representing the finance industry on the global level, also merits mentioning. IIF published a detailed report describing possible regtech products and services combined with proposals for implementation.[18] Since the IIF is not a public body, however, its activities can only show possible pathways to the regulators.

3.2 U.K.

The most active player in regtech has been the U.K. Financial Conduct Authority (FCA). As mentioned, the FCA coined the term regtech in 2015 and published a "Call for Input," inviting market participants to describe their activities and objectives. The success of the consultation was noteworthy in respect of the number of participating enterprises as well as of the involved high-ranking officers.

On July 20, 2016, the FCA published a Feedback Statement summarizing the contents of the submissions.[19] In addition, the FCA formulated four strategic objectives, namely:

- **Efficiency in data exchange:** cloud services or other online platforms should be used by financial intermediaries to increase efficiency at the interface between enterprises and supervisory authorities.
- **Standardized and automated data processing:** regulations should be machine-readable or even interactive (robo Handbook) in order to decrease the costs, particularly in case of ongoing regulatory reforms; the API needs to be designed in a way that systems can interact with each other without data losses or other problems.

[15] Financial Stability Board, 2017, "Financial stability implications from FinTech – supervisory and regulatory issues that merit authorities' attention," June 27, http://bit.ly/2seQzE4; see also FSB/IMF, 2016, "The financial crisis and information gaps: second phase of the G-20 data gaps initiative. (DGI-2) first progress report," http://bit.ly/2lgINGL.
[16] FATF FinTech and Regtech Forum 2017, "Guiding principles (San José principles)," http://bit.ly/2xRIJ4d. See also BIZ: FX Global Code, "A set of global principles of good practice in the foreign exchange market," http://bit.ly/2tiqunG.
[17] IOSCO research report on financial technologies (FinTech), February 2017, http://bit.ly/2kZVzlc.
[18] Institute of International Finance (IIF), 2016, "Regtech in financial services: technology solutions for compliance and reporting", Washington DC, March 22, http://bit.ly/2xROajK.
[19] Financial Conduct Authority (FCA), 2016, "Feedback statement, call for input on supporting the development and adopters of Regtech, FS16/4", July 20, http://bit.ly/2x9z4m9.

- **Systematic data analysis and data prognosis:** analytics solutions should allow the creation of so-called "data lakes." As a consequence, the interpretation of extremely wide ranging and unstructured data would become more accurate.
- **Innovative approaches for compliance processes:** the use of DLT for the storage, verification, and encryption of data should be promoted, mainly in the context of the KYC principle and the anti-money laundering (AML) regulations.

In addition, the FCA invites market participants to pay much more attention to regtech products and services and to invest funds into the development of new processes in this field. Furthermore, the creation of machine-readable formats for the "Regulatory Handbook" is proposed [Contratto (2016)].

3.3 Hong Kong

The Hong Kong Monetary Authority (HKMA) published a report about innovative finance technologies in February 2016.[20] The main focus of this report was defining the conditions for Hong Kong to become a "fintech hub," however, the advantages of the use of regtech are also mentioned.[21] In particular, market participants are invited to more thoroughly realize the potential of regtech by cooperating with technology enterprises.

3.4 Singapore

The Monetary Authority of Singapore (MAS) organized a "fintech festival" in November 2016, which also tackled regtech issues. In the meantime, the fintech Guidelines of the MAS are published.[22]

3.5. Australia

The Australian Security and Investment Authority (ASIC) established a regtech team in June 2016.[23] The task of this team consists of cooperating with other organizations related to the development of new regtech software.

3.6 Switzerland

In March 2016, a member of the House of Representatives filed a motion under the title "Improvement of the Digitization in the Regulation (regtech)"; therein, the Federal Council is asked to analyze the possibilities for broadening the scope of application of regtech initiatives.[24] The Federal Council has agreed to consider the related questions within the realization of the parliamentary motion no. 15.4086 "Für einen wettbewerbsfähigen Finanzplatz im Bereich neuer Finanztechnologien" (competitive finance market in the area of financial technologies). Additionally, a report regarding the framework requirements for the digitization of the economy has been published.[25]

On February 1st, 2017, the Federal Council initiated the consultation on amendments to the Banking Act (BankA) and the Banking Ordinance (BankO) in the fintech area. The proposed amendments should ensure that barriers to the market entry of fintech firms are reduced by separately regulating the firms that provide services outside the normal banking business according to their risk potential.[26] The three deregulation elements are: (i) the holding period for the acceptance of funds for settlement purposes (Art. 5 para. 3 lit. c of the BankO) is extended (60 instead 7 days); (ii) an innovation area (sandbox) is introduced in which the acceptance of public funds up to CHF 1 million is not classified as operating a banking business; as a consequence, this activity is exempted from authorization; and (iii) simplified authorization and operating requirements compared to the current banking license apply in the areas of accounting, auditing, and deposit protection for companies that accept public funds of up to a maximum of CHF 100 million but do not operate in the lending business.

In July 2017, the Federal Council adopted the amendment of the Banking Ordinance, which entered into force on August 1st, 2017.[27] The other amendment that concerns the Banking Act was already addressed by the parliament in the context of the deliberations on the Financial Services Act (FinSA) and the Financial Institutions Act (FinIA). The new authorization category will be created in the BankA for companies that accept public funds of up to a maximum of CHF 100 million but do not invest funds or pay interest on funds. For this new fintech category, simplified authorization and operating requirements in the areas of accounting, auditing, and deposit protection will be applicable.

---

[20] Hong Kong Monetary Authority (HKMA), 2016, "Report of the Steering Group on Financial Technologies," February 26, http://bit.ly/1pmsWmR.
[21] "Moreover, regulators themselves may also benefit from FinTech. For instance, regulatory authorities in most financial centers around the world are increasing the deployment of technology to complement existing regulatory processes and facilitate more effective risk identification, risk weighting, surveillance, and data analytics, commonly referred to as "Regtech". Regtech results in opportunities for the FinTech sector to provide data standardization, collection, visualization, and analytics solutions to both regulators and regulates." HKMA (2016), 20.
[22] Monetary Authority of Singapore (MAS), 2016, "Regulatory sandbox guidelines," November, http://bit.ly/2gDPTAm.
[23] Australian Securities and Investment Commission, 2017, "Response to feedback on REP 523 ASIC's Innovation Hub and our approach to regulatory technology," September, http://bit.ly/2ytWW4l.
[24] http://bit.ly/2xU2Pva.
[25] Bericht über die zentralen Rahmenbedingungen für die digitale Wirtschaft, January 11, 2017, http://bit.ly/2jDjSeE.
[26] See Federal Council, media release, February 1, 2017, http://bit.ly/2wsye49.
[27] See Federal Council, media release, July 5, 2017, http://bit.ly/2xJHqVQ.

Looking from a general regtech perspective, several possibilities to exchange information in a digital way have already been introduced within the financial services sector [Contratto (2016), no. 34]:

- License procedures for collective investment schemes.
- Financial reporting for collective investment schemes.
- Delivery of data by banks to the Swiss National Bank.
- Self-declarations in the context of anti-money laundering legislation.
- Notification obligations of securities' offerors to the SIX Swiss Exchange.
- SIX Swiss Exchange approval for the trading of capital market products.

Apart from the mentioned examples, the automated transaction surveillance and the name-matching systems are accepted digital measures helping to comply with regulatory submission tasks.

4. LEGAL CHALLENGES TO REGTECH

The implementation of regtech products and services creates a number of legal challenges, which must be met in order to achieve a successful implementation.

4.1. System resilience and risk management

System resilience is a key challenge in the regtech environment. Major emphasis must be placed on the areas of data protection and cyber risks, having consequences for IT architecture, the design of IT systems, and the governance/compliance in this field.[28] The implementation of Standard 239 of the Basel Committee for Banking Supervision (BCBS) related to the principles for effective risk data accretion and risk reporting (January 2015) has demonstrated the respective difficulties [Contratto (2016), no. 39].

Looking from the perspective of the regulators, a failure to develop the IT capabilities enabling the use of the data provided in response to reporting requirements would severally impact the achievement of the policy objectives of such reporting requirements [Arner et al. (2017), 399]. In other words, the regulators themselves should provide an example of a forward-looking implementation of IT-enabled systems. Reporting must be matched with analytical tools, as good practice has already shown in the securities markets.

As far as business entities are concerned, system stability, as well as system resilience, require the implementation of risk management and compliance measures. The confidentiality, integrity, and availability of data must be guaranteed not only for business operations but also for the fulfillment of the regtech conditions.[29] The internal risk management manual has to describe the implementation of digital transactions and the appropriate procedures for the protection of data (authentication, identification). A respective IT governance framework could be designed as follows:[30]

OBJECTIVES OF IT GOVERNANCE
- Development of environment
- Security of IT infrastructure/ services
- Regulatory context
- Plan for extraordinary events

IMPLEMENTATION OF MEASURES
- Program administration
- Education
- Documentation
- Human and financial resources

COMPLIANCE
- Monitoring and surveillance
- Reporting
- Improvements (after incidents)

Cyber risk prevention has been another topic for financial intermediaries for a number of years. With the development and implementation of regtech, the challenges will increase further and the regulatory risk management concepts will need to be adapted to the new digital applications [Larsen and Gilani (2017), 27/28].

4.2 Coordination and cooperation requirements

The supervisory authorities of the U.K. (Financial Conduct Authority, Payment Systems Regulator, Prudential Regulation Authority, Bank of England) are advised by way of the so-called Productivity Plan 2015 of the British Government to closely cooperate in respect of the development of regtech.[31] Even budgetary measures are introduced or proposed for implementation.

[28] See also Huber, R., 2017, "Kann Regtech die Finanzindsutrie aus der regulatorischen Krise befreien?" Weblaw Jusletter May 18, no. 8.
[29] Weber, R. H., 2017, "Systemstabilität: neue Herausforderungen durch die Digitalisierung der Geschäftsmodelle," in Festschrift für Hans Caspar von der Crone, Zurich, 405, 410.
[30] Weber, R. H., 2017, "IT-governance: unverzichtbar für jedes Unternehmen, schulthess manager handbuch 2017," Zurich, 37, 38/9.
[31] HM Treasury, Budget 2015, March 2015 (http://bit.ly/2x8T8d4), 53: "[... the FCA, working with the PRA, will also identify ways to support the adoption of new technologies to facilitate the delivery of regulatory requirements – so called Regtech]."

Such forms of coordination would equally be reasonable for other national regulators and even for cross-border regulatory activities [Contratto (2016), no. 37]. In the medium-term, it should also be considered to establish certification tools for regtech applications and regtech providers (possibly similar than in data protection law); thereby, quality standards are likely to become higher.

Improved coordination and cooperation[32] can equally be achieved through a regular exchange of information among the involved interest groups and commercial entities; not only banks but also multinational enterprises [Contratto (2016), no. 38]. Since businesses often act across multiple markets, IT-based systems must be implemented that not only cover specific requirements of national law but also of global standards. Obviously, competition law sets some limits to the information exchange; however, networking in form of roundtables, discussion fora, and conferences should not be critical.

5. OUTLOOK

In view of the digitization of many parts of the economy, regtech will certainly gain importance and constitute a growing challenge for private organizations as well as for regulators. As far as the businesses are concerned, the new technological environment requires the implementation of better risk management and IT compliance systems; attention must be diverted from traditional regulatory topics to the needs of a new technological environment. The developments in distributed ledger technologies (blockchain) as used in securities markets are a good example.

Most likely, the change will even be more "dramatic" for the regulators and the supervisory authorities. A paradigm shift is expected to occur in financial regulation. Among others, the well-known principle of KYC might be complemented by a new principle of KYD.[33] This new perspective leads to more data-driven regulations and compliance models. An interesting example is the so-called "India Stack," a joint private-public project encompassing four level for a fintech transformation.

[32] Huber (2017), no. 7; see also Contratto, F., 2017, „Technologie und Finanzmarktregulierung: Narrative von Interdependenz und Co-Evolution," in Festschrift für Hans Caspar von der Crone, Zurich, 421, 430/1.
[33] Arner, D. W., J. Barberis, and R. P. Buckley, 2016, "The emergence of Regtech 2.0: from know your customer to know your data," Journal of Financial Transformation 44, 79 et seq.

Elements of such a new regulatory approach, which should be considered in depth, could encompass the following phenomena [Arner et al. (2017), 405/06]:

- Regulatory policy modelling: agent-based modelling as an emerging technique.
- Reporting standards: common compliance standards across multiple jurisdictions.
- Systemic risk tools: mathematical tools developed in cooperation with central banks.
- Harmonization: integration of monitoring systems beyond varied regulatory demands.
- Uniform compliance tools: development of a suite of open source compliance tools.
- Collaboration and selected data sharing: improved cooperation of regulators through transformative processes.

The challenges evolve rapidly in the technological context. Looking forward, therefore, regtech should be built into a reconceptualized (financial) regulatory regime.[34] The paradigm shift based on a holistic approach will have to put more emphasis on the nexus of data and digital identity. Data needs to be gathered globally on a real-time basis and processed in a coherent way in order to meet the cross-border needs of businesses and regulators.

[34] Arner et al. (2017), 376 and 414. See also Baxter, L. G., 2016, "Adaptive financial regulation and Regtech: a concept article on realistic protection for victims of bank failures,", Duke Law Journal 66, 567 et seq.

Bridging the gap between investment banking infrastructure and distributed ledgers

MARTIN WALKER | Banking & Finance Director, Center for Evidence-Based Management
ANTON SEMENOV | Principal Business Analyst, Commerzbank AG

ABSTRACT

Enthusiasm for blockchain, or the broader family of distributed ledger technologies (DLT), within capital markets is now into its third year. The enthusiasm has manifested itself in numerous pilots, proofs of concept, fintech startups and industry collaborations. Within investment banks, large broker-dealers, and many "buy-side" firms, the enthusiasm has in large part been driven by a combination of heavy demands on IT departments and considerable pressures to cut costs. This paper argues that there is a key factor that has prevented the delivery of any significant working systems within these enterprises; a general tendency to try to fit a solution (blockchain and its derivatives) to problems, rather than trying to understand problems and find the appropriate solutions.

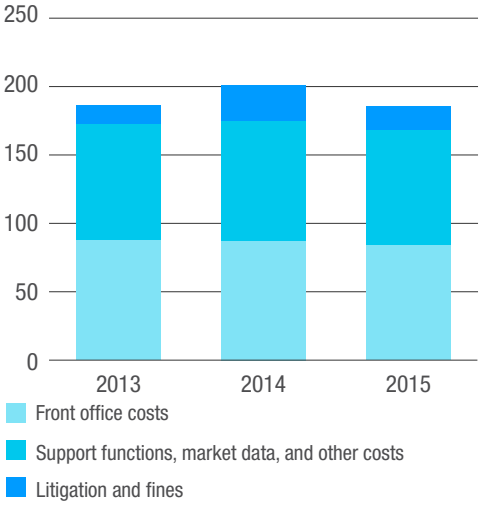
Relative analysis of firm infrastructure suggests that the root causes of most problems are not technical but are human in nature; being related to incentives, culture and organizational structure. The analysis demonstrates that trade processing data and business logic are highly distributed but frequently highly inconsistent. A model is proposed (drawing on many of the concepts of DLT) to create both transparency of issues and mechanisms for the propagation of consistent business logic and consistent data models. This aims to use DLT based techniques to deal with fundamentally human problems by the introduction of the appropriate feedback loops for management decision making. Understanding the nature of problems and the effectiveness of changes would allow genuinely evidence-based management decision making. A technologically driven, human transformation that could act as a lever for unravelling organizational complexity.

1. INTRODUCTION

Specialist investment banks and the "markets" divisions of the universal banks execute millions of trades each day, with total notional values running into trillions of dollars. The banks' main trading partners include fund managers, pension funds, hedge funds, large non-financial corporations, and other banks.

This vast amount of trading activity is not just dependent on traders. There are complex IT infrastructures inside each bank and large numbers of support staff in critical functions, such as risk, finance, and operations. Research by the Boston Consulting Group estimated that the total cost of IT and other support functions within major financial institutions that supported trading in financial products was around U.S.$83.9 billion in 2015. This is a number that is both incredibly large and remarkably resistant to the cost reduction efforts of banks, as shown in Figure 1.

Figure 1: Investment banking operating expenses (U.S.$ bln)



Source: Global capital markets 2016: the value migration, BCG Perspectives, http://on.bcg.com/2tmMjxM

Supporting all this trading activity (and generally reflected in the costs discussed) are many external organizations (and their infrastructure), including brokers, market data providers, central securities depositories (CSD), trading platforms, exchanges, matching platforms, and clearing houses. The "front-to-back" processing of a trade will typically involve many systems (internal and external to the bank) and parties.

Investment banking, which also includes businesses that are not explicitly related to trading but often result in trades, such as equity issuance, debt issuance, and mergers and acquisitions (M&A), has always been a highly cyclical business, with profits dramatically rising and falling in line with market activity. Historically, one of the key shock absorbers of this volatile business has been staffing costs. Good times meant large bonuses, bad times meant no bonuses and staff cuts. However, since the Great Financial Crisis (GFC) of 2007-2009, trading activity has generally proven to be subdued, normal operating costs have become harder to reduce (in spite of significant reductions in the size of the front office), and control and compliance costs, along with the on-going fines and litigation costs, have grown substantially.[1] Additionally, regulations limiting the banks' trading activities, requirements for additional capital, and restrictive targets for liquidity and leverage have severely reduced the scope for increasing revenues.

It is not surprising that in such an environment investment banks are more focused than ever on reducing operating costs. However, the stickiness of costs has meant that they have had to consider a range of new, and not so new, ideas, such as greater use of off-shoring and outsourcing, mutualization of business functions, digitalization, and the application of distributed ledger technology (DLT).

Given the amount of interest expressed by investment banks in DLT over the past few years it is worthwhile to consider its implications. And, while it is true that many banks, consortia, and fintech companies have undertaken successful trials of the technology, the question remains as to whether it will be able help to reduce the aforementioned costs.

This paper argues that a pragmatic, hybrid approach to applying distributed ledger-like technologies can help reduce the cost of processing trades in financial institutions. However, for that to happen, it does need to be applied in ways that might be quite different to those currently being considered. This is because we believe that even if DLT does take off on a large scale in investment banking it will be part of a mixed environment of centralized and distributed systems for a long period. A pragmatic, hybrid approach would also make it easier for production quality DLT solutions to meet the banking sector's strict requirements in relation to procurement, security, and data privacy.

[1] Fines and litigation costs have been so large in both absolute and relative terms they have formed one of the main drags on profitability.

To create value for banks in the short-term, DLT needs to be applied in a way that is not dependent on a "big bang" replacement of infrastructure. Even if significant DLT-based applications are rolled out, they will need to be capable of integration with a great deal of existing infrastructure. Those who have worked in investment banking IT departments know that, in spite of the glamour of implementing new trading algorithms and low latency trading infrastructure, the bulk of the work comes down to the unglamorous but critically important job of integrating systems – "the plumbing."

A hybrid approach that includes elements of "big data"/analytics, as well as "nudges" of behavioral economics, can help banks tackle one of the major challenges they face implementing large scale change, namely complexity. Complexity makes organizations (and systems) hard to measure, understand, and consequently change. DLT combined with "big data" has the potential to introduce the transparency and feedback loops that are missing in complex organizations. With those in place, there is the potential to "nudge" organizations towards the standardization and behavioral change, which would ultimately reduce costs and operational risk.

2. WHAT DO WE MEAN BY TRADE PROCESSING?

When considering complexity within an investment bank it is very easy to lose sight of what all those systems and departments, spread across so many locations, are ultimately trying to achieve. While it is important not to trivialize any of those functions, it is possible to look at trade processing in a way that makes things much clearer.

At its most basic level, trade processing within a bank consists of taking a small set of inputs from external systems or the physical world (i.e., traders) to generate a very large number of outputs. These are the outputs needed by a bank to answer fundamental questions like, "How much money is being made?", "What risks are being taken?", "What trades could or should we do next?", etc. There are also outputs that are needed to meet the requirements of regulators, accounting standards bodies, and clients.

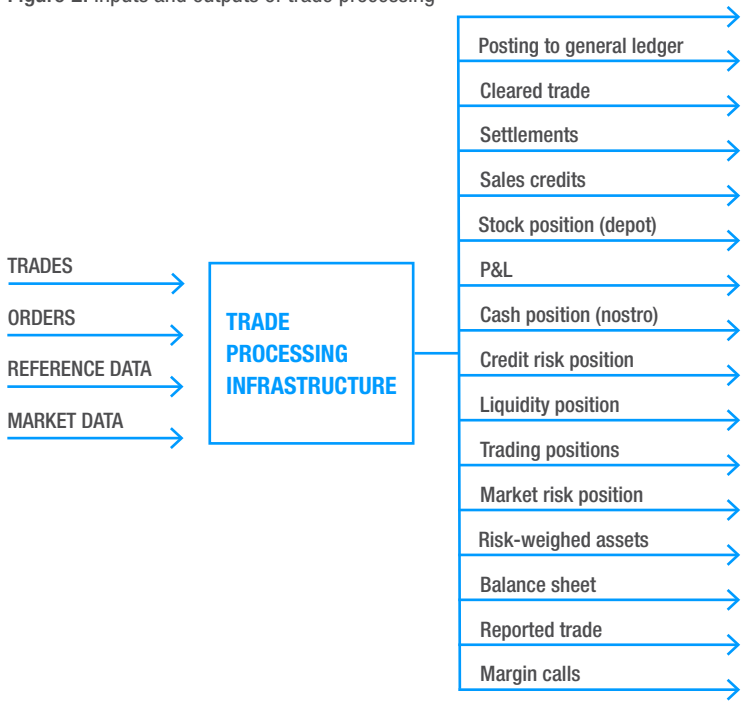
This does not mean that the optimal infrastructure would consist of simply creating the box labelled "Trade Pocessing Infrastructure" in figure 2 as a single "smart contract."[2] There would be nothing particularly simple about turning all the required business logic into a DLT-style "smart contract" and it would be a very complex (and hard to maintain) piece of code.

2.1 A model infrastructure

Another key point to understand about the trading infrastructures of most investment banks is that broadly speaking they work. Trades get settled, credit limit utilizations updated, and postings are made to the general ledger. The extent to which that infrastructure is cost effective, flexible, and controlled varies enormously between banks. In spite of the problems, which are to some extent shared across all banks, it is possible to draw a picture of what "good infrastructure" looks like, using existing technologies and techniques.

Looking across the trade processing stacks of most banks, for each main asset class, a pattern emerges. A low latency, connectivity layer connects the trading businesses with exchanges, ECNs (electronic communication networks), SEFs (swaps execution facilities), and the firm's own external facing trading platforms. This needs to be fast and it needs to be highly resilient. Bursts of tens of thousands of trades may hit this infrastructure over the course of seconds.

Figure 2: Inputs and outputs of trade processing



[2] A reasonably commonly accepted definition of a "smart contract" is a code that can run autonomously to enforce and execute the terms of a contract. However, common usage of the term may vary between any programs running in a distributed ledger to genuinely independent objects that could theoretically have their own legal personality.

Those trades need to feed into the bank in as frictionless a manner as possible so that they can be quickly credit checked, executed, hedged, and used to update positions and bids and offers. Speed is essential because a slow moving bank will find itself on the wrong end of trades with faster moving rivals or the even faster algorithmic hedge funds. Resilience and lack of friction is just as important as speed. There is no point being able to execute large numbers of trades and orders if they get "stuck in pipes" due to connectivity or static data issues.

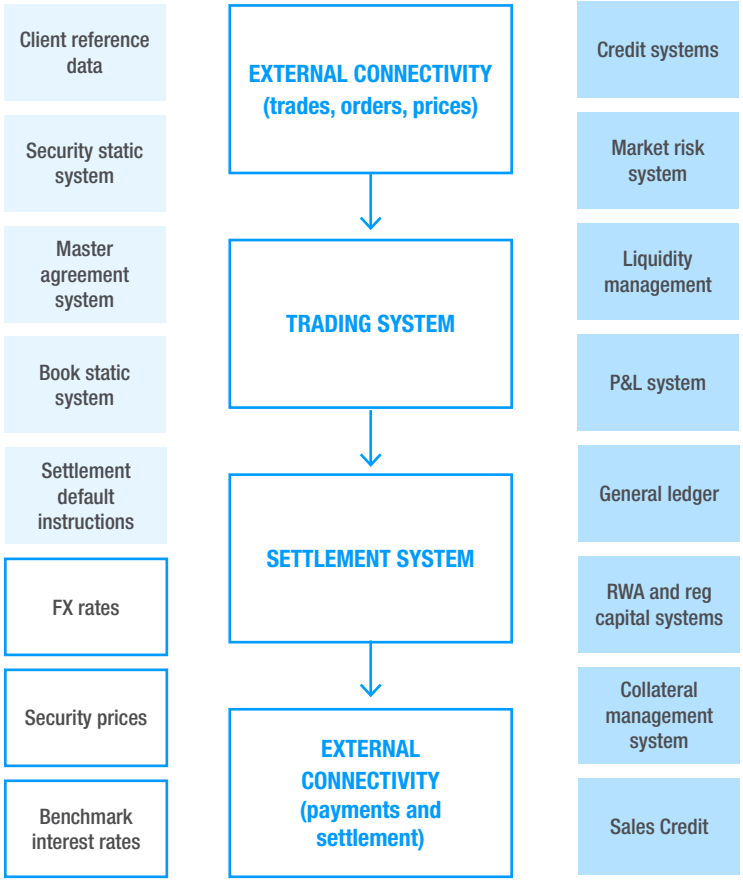
While some elements of this low latency infrastructure can be bought off the shelf, some banks have had very clever technologists building sophisticated infrastructure in this area. Highly paid IT professionals have also been hired to build this type of infrastructure in hedge funds, and in some cases they are far more sophisticated than the banks they trade with.

Trades and orders are captured in this layer and are typically fed into a trading system.[3] Historically, the trading system[3] was the place were trades were booked, positions managed, and risk and P&L calculated and viewed, i.e., where traders did their trading. Today, the "trading system" is turning into something of a misnomer with ever greater proportions of trades being executed on external platforms, more trading decisions being taken by automated processes, and risk and P&L being viewed on a cross-asset class basis.

There is still, however, the need for a central view of risk and P&L. There are also the more complex, the less liquid, and the "voice" trades; all of which need people and a trading system. Good trading infrastructure below the low latency layer may not need to work at quite such a frantic pace but it still requires to be well integrated with other infrastructure, to have good quality reference data, and the right business logic. Both internal or vendor systems in this space are often very mature (in a good way) and sophisticated. A bank can "mess up" the integration and configuration of even the best systems, but good systems combined with good integration can create very smoothly running infrastructure in this area.

It would be easy to assume that trades are then fed from the trading system to the settlement system where trades are then "settled." However, a better way of describing a settlement system (at least in a generic, cross-product way) is a place where trades are made ready for settlement. Cash flows and stock movements may be generated from the trades and trade events received from trading systems. Records of stock at depos and cash in Nostro accounts may be updated and various exceptions resolved by the operations staff. When everything is ready, the instructions to move cash or securities are communicated to the outside world (or in some cases other parts of the bank).

Figure 3: Core trade flow and related systems



The settlement infrastructure (and often closely related infrastructure for post-trade confirmation and matching) is typically the place where the noise of bad reference data and mis-booked trades becomes apparent. However, if reference data feeds are good, everything upstream is integrated well, and the traders show a disciplined approach to booking their trades, the operational systems can work relatively smoothly. Despite that, counterparties can still inflict pain on the best operations department. For example, they can incorrectly book their side of the trade, they can demand strange quirks in the post trade processing, or simply be unresponsive when errors are found.[4]

[3] Trades, quotes, orders, and availability are also generally generated in this layer to feed to external platforms.
[4] Examples include ad hoc changes to netted settlements, requests for non-standard information to be added to confirmation, third party payments of the back of an FX transaction

However, a well implemented post-trade infrastructure should be able to tell you the true cost of dealing with troublesome counterparties and provide the data to encourage better behavior.

Settlement systems share many of the same attributes as the trading systems. There are many mature, high quality systems available from vendors, but even the best of these can be perform poorly due to poor configuration and integration. Some banks build and maintain their own settlement infrastructure, which may reflect the strength of having mature high volume system, though in some cases the persistence of in-house systems may simply reflect the difficulty of replacing them.

At the bottom of the trade processing layer are the systems that connect the bank to the places that will actually move the cash and/or securities, the custodians, CSDs, nostro banks, etc.

Interacting with the core trade processing systems are the systems owned by the major support functions, notably those associated with the risk and finance departments. While the core trade processing systems are frequently specific to an asset class, those supporting other functions are generally cross-asset class. P&Ls[5] need to be generated for the bank as a whole, as well as at the level of the trading desk or a book. Risk related to a particular counterparty needs to be viewed across all business lines. Although we have discussed at some length the trading and settlement systems, the costs of these systems can be much higher than the core trade processing systems and they may contain a great deal of business logic.

Risk, finance, and related systems have exactly the same dependencies on good integration and good reference data. They work extremely well in some banks but not so well in others, requiring significant manual intervention in core business processes. Perhaps the alarming consequence of poorly plumbed-in risk and finance infrastructure is that problems can remain obscured for much longer, as many banks discovered during the GFC.

Linked to the main trade processing systems, and in many cases the systems of support functions, are (or should be) centralized systems providing the reference and market data needed for calculations and trade enrichment.

Acting as the backbone for most of this infrastructure is a messaging layer that allows communication between all these systems in near real-time and with guaranteed delivery.[6] However, a significant degree of data is still exchanged as part of end-of-day batch processes, using a wide variety of methods. There may be SFTP (secure file transfer protocol), direct database connections, or even the use of the messaging layer. Whatever the methods used for communication, good infrastructure ensures the data reaches its destination in a timely manner without being lost or mutated.

Other factors that should help the creation of smoothly running infrastructure are standardized messaging formats, such as FpML[7] (for trades and the trade events), and a huge range of off-the-shelf software packages. These may be product specific, function specific, or a combination. Adding all these elements together, using consistent market and static data, and making sure that all traded products are fully supported by all the relevant systems, with messages that mean what they say, can create an efficient and low friction infrastructure.

2.2 The problem of trade processing

Investment banks today can process vastly more trades, more quickly, and at lower cost than in previous decades. Products that in previous decades could not even be imagined are traded and processed on a daily basis. Banks would particularly like to reduce their technology costs since they are large in absolute terms and as a proportion of revenues. However, other sectors accept the need to spend money on technology. It may cost car manufacturers around a billion dollars to build a new plant but they do not simply try to wish away the costs of doing business.

The difference in investment banking is the resources devoted to, frequently fruitless, attempts to improve the infrastructure. Generally, the bigger and more ambitious the project, the greater the risk of failure. Front-to-back re-engineering, “simplification”, front-to-back systems, horizontally organized functional systems, and clever (but expensive) middleware programs are all approaches that have been tried multiple times across the banks with very varying degrees of success. There have also been many complete failures. The other

[5] Which in a non-banking environment equates to revenues rather than profits
[6] The messaging system uses a built-in data store to persist messages. It does not guarantee that the recipient can successfully process the message.
[7] Financial Products Mark-Up Language

problem is the "noise" that develops due to progressive deviations from the model infrastructure described. Sources of this noise include:

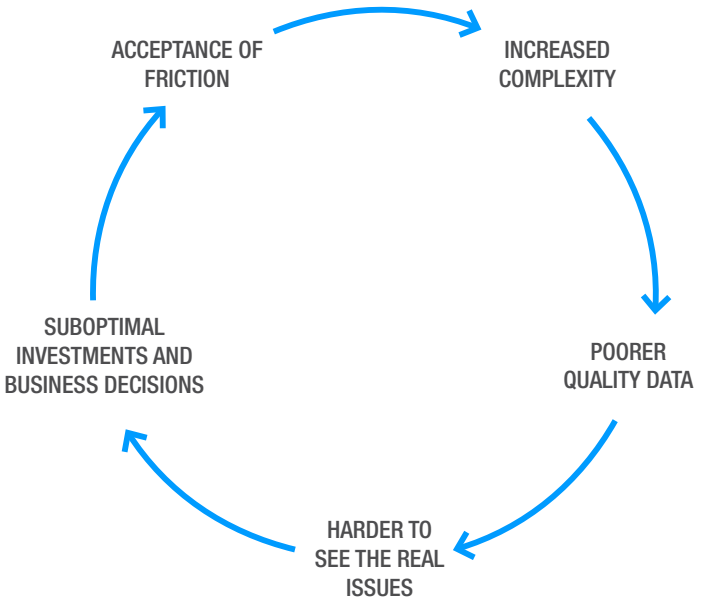
- Problems in the quality and completeness of messaging between systems. Some front office IT staff use the terrible phrase "fire and forget;" send the trade downstream but do not worry about whether anyone can make sense of it or not.
- Lack of reference data systems or failure to connect all relevant systems to those reference data systems.
- Poor trade booking due to the errors of front office staff combined with a failure to encourage better practices.
- Trades and structures that are booked in one systems but not understood by the systems they are fed to (even if the trades are messaged correctly).
- Bespoke processes for clients.
- Significant volumes of manually booked "voice" trades that are inherently prone to error.
- Manual and/or paper-based confirmation and matching processes.
- The mutation of "standard" messaging formats to make up for problems in other parts of the system infrastructure.

There are large variations in the degree of noise between banks, or within banks, between different businesses, or regions; a theme almost every investment banker would recognize. To put it more simply, superficially similar system infrastructures can have widely varying costs and levels of operational risk because some are not wired together properly, lack support for the products traded, or miss the relevant reference and trade data sourced from golden sources.

"Noise," or "friction," has other indirect costs in addition to the labor costs of paying people to fix problems; it needs more layers of control. In a fragmented, "noisy" infrastructure, those layers of bolt-on (rather than integrated) controls can become another source of noise and error. All the resulting complexity becomes progressively more expensive to run (and to change) because the quality of data required to make the right decisions deteriorates as it becomes more dependent on people and interpretation.

Regulators, the technical press, and even senior bankers have grown highly critical of the state of investment banking infrastructure. Criticism extending beyond cost to operational risk, quality of data produced, and flexibility to deal with changing markets and regulations. One of the most recent tests of the banks' responsiveness to change were the trade reporting requirements under the Dodd-Frank Act (DFA) and the European Market Infrastructure Regulation (EMIR). These requirements were superficially straightforward:

Figure 4: The complexity cycle



to report both the trade economics of OTC derivatives in near real-time and post-trade data by the end of the day. The exercise proved problematic, if not traumatic, for most banks with costs ranging from the tens to hundreds of millions of dollars per bank.

The complexity discussed is not necessarily any specific person's fault. Much of it has been the result of:

- Mergers and acquisitions by the banks and failure to completely integrate infrastructure and businesses.
- Two decades of breakneck financial innovation, including the creation of hybrid products and structures.
- Decades of largely autonomous business units and legal entities making decisions that were right at the time for their entity or business but ultimately wrong for the organization.
- The development of capital markets into a collection of genuinely global businesses.
- The errors that can arise every time a trade is translated from one system's data structure into a messaging format then translated again into the receiving system's data structure.

2.3 The limits of centralization and opportunity for DLT

One of the areas where the greatest progress has been made in trade processing in recent decades has been the efforts to increase standardization and centralization[8] of processing. The huge increases in volumes and variety of products would not have been possible without **centralized** services, such as trading ECNs, the SWIFT network, CSDs, Euroclear and CLS, and **standardization** of product definitions, legal documentation, and message types through the work or organizations such as ISDA, SWIFT, and ISLA.

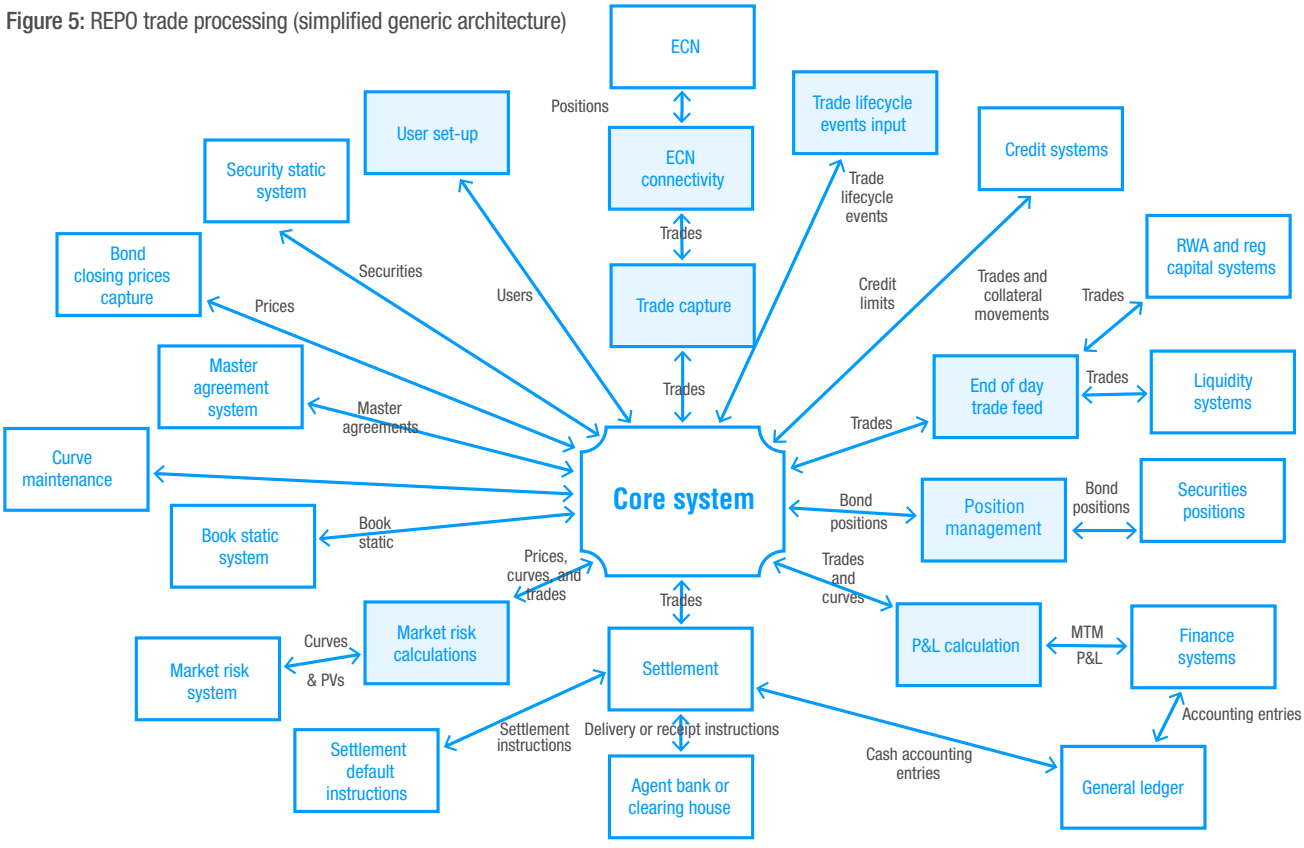
In recent years, there has been increased desire to extend this approach to the cost challenge. Banks have been more open than ever to the idea of mutualization of their system functionality in areas that they do not consider to be differentiating. However, creating mutualized utilities has not proven easy. Infrastructure in the middle parts of the trade lifecycle have proven particularly hard to mutualize because the systems in those areas have to deal with the most friction-driven complexity.

Even for better built infrastructure, basic complexity theory kicks in because of the greater number of connections between systems in the middle of the trade processing lifecycle. To use the technical term, complexity is a function of the number of "edges" and "nodes" as shown in Figure 5.

Ideally a technology would be available that:

- Has the same magic ingredient of the best centralized systems, i.e., standardized data models and business logic.
- Deals with the basic drivers of complexity, i.e., reduces the number of edges.
- Supports the mutualization of non-differentiating processing between banks.
- Can be combined with analytics software to make it easier to measure both problems and the impact of changes. Providing a tool for better management decision making.
- Potentially, simplifies the settlement process and reduces the time take for settlement cycles.

Figure 5: REPO trade processing (simplified generic architecture)



[8] Not something you may have anticipated reading in an article about distributed ledgers

Many people would argue that DLT is that technology. However, the question is how much of that is really plausible in the short term?

2.4 The barriers to ledger nirvana in trade processing

In "ledger nirvana," market counterparties use consistent sets of trade data with smart contracts that apply consistent business logic to produce the various outputs required to operate a trading business (Figure 2). However, in spite of many proofs of concept in various aspects of capital markets and the major investments made by some DLT-related companies, there are a number of obstacles that need to be worked through to in order facilitate more widespread of adoption of DLT in markets. **None are insurmountable but all could take considerable time and effort.**

A distributed ledger-based solution to trade settlement needs to be tangibly better in terms of cost, control, security, and resilience than financial market infrastructures already in place; much of which works remarkably well, such as the major CSDs and CLS Bank in the FX market.

In ledger nirvana, settlement infrastructure is typically based on the assumption of cash and securities on ledger, i.e., cash or securities legally exist on the ledger or have a "tokenized" representation of assets that enjoy the same degree of legal certainty and settlement finality as the primary record of the assets.[9] Some progress has been made in this area with Overstock's issuance of shares on its TO platform, which includes the use of a private ledger performing core processing with all transactions ultimately being recorded on the Bitcoin blockchain. This represents a small step forward legally and technically but is still a long way from making a significant impact.

Genuine "cash on ledger" is even more problematic. Fundamentally, a real-world fiat currency needs to be a ledger version of central bank reserves, which require the cooperation of central banks and in some jurisdictions legal changes, or commercial bank created money. Commercial bank money would be the direct economic and legal equivalent of the money that is represented today as positive balances in customers' bank accounts. However, commercial bank issued "ledger money" would still have credit risk against the issuing bank and needs a mechanism (equivalent to the central bank clearing systems) to control the credit risk that builds up between clearing banks as funds are transferred.

For securities, such as government bonds and more liquid equities, mechanisms would be needed to make it possible for them to be provided as collateral between banks, to CCPs, and central banks through their repo process, in addition to simply being bought and sold.

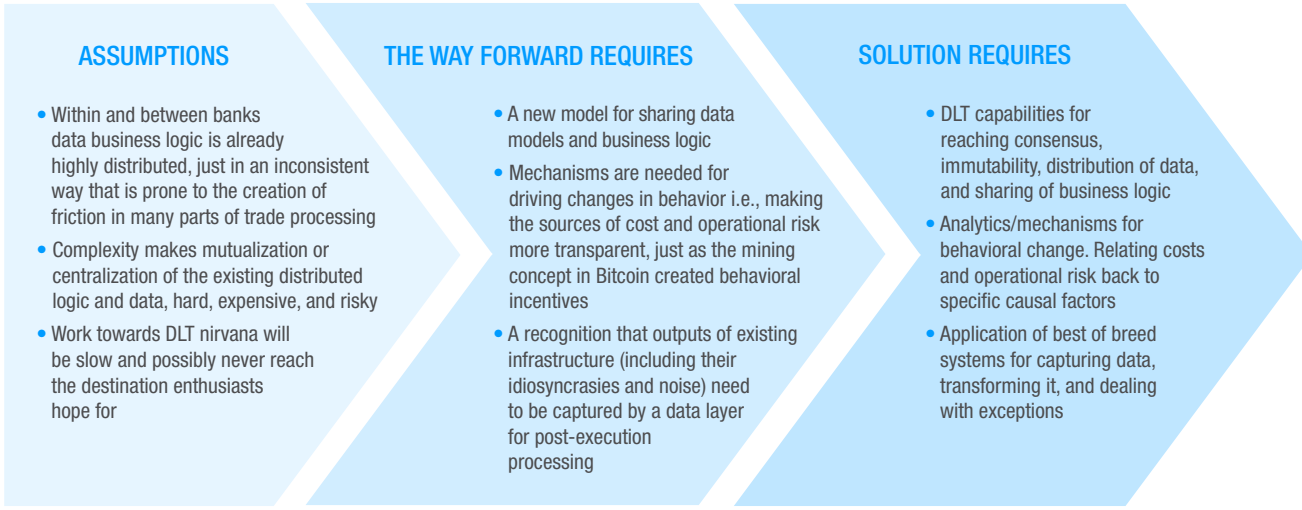
Smart contracts that implement the mechanics of a financial product, such as the work done by Axoni on equity swaps or Barclays and R3 on interest rate swaps, represent a step forward in supporting the trade lifecycle of derivatives trades using DLT. However, there are many systems within the trading infrastructure of an investment bank that execute, enrich, process, and aggregate trades and trade events. A smart contract that performs the basic mechanics would still need to interact with credit risk, market risk, liquidity management, position viewing, P&L calculation/aggregation, regulatory report, derivative clearing, sales credits, and many other systems. What is frequently forgotten is that simply having a ledger of trades does not remove the need for a general ledger, frequently the most complicated and expensive system in a bank. A typical general ledger system is not just a list of transactions. It is also a list of accounting rules and policies that are applied to the transactions, often requiring the support and judgments of a large finance department.

"A smart contract that performs the basic mechanics would still need to interact with credit risk, market risk, liquidity management, position viewing, P&L calculation/aggregation, regulatory report, derivative clearing, sales credits, and many other systems."

In many markets, such as spot FX, futures, cash equities, and the more liquid bond issues, the majority of trading (including much order processing) takes place at very high speed using very expensive and sophisticated infrastructure. It can be argued that this speed does not add significant value to society or the economy but it is the reality of how many markets operate today. There would be great resistance by the markets to any attempt to slow down trading to allow DLT, which is inherently slower, to replace the current pre-trade

[9] Tokenization is still largely a concept rather than legally and regulatory verified reality.

Figure 6: From assumptions to solutions



infrastructure. A further obstacle is that the post-trade processing costs of electronically executed trades are considerably lower than for the more traditional (and error prone) voice trading. This means that DLT solutions need to be significantly better post-trade than systems and processes that deal with relatively trouble free "e-trading."

Ledger nirvana would also make many proposed distributed ledger-based systems fall within the scope of the Bank for International Settlements' (BIS) "Principles for financial market infrastructures."[10] These principles are incorporated in law in most jurisdictions and are justifiably demanding. They represent a high hurdle for DLT to clear.

Overlapping with the BIS principles are the banks' own requirements for high volume processing, resilience, and security. Just as the existence of a DL does not automatically remove the need for a general ledger, the use of cryptographic techniques does not make a system more secure from a bank's perspective.

In ledger nirvana, the trade is the settlement. A trade is booked and value exchanged. However, this creates significant problems for today's business models, which cannot simply be wished away by the DLT enthusiast. Most of capital markets works implicitly on time delays. Huge daily volumes are traded and processed but a market maker only needs to be flat (in most markets) by the end of the day. The settlements teams only need to transfer the net settlement amounts at the end of the settlement cycle. In the world of "trade equals settlement," a market marker can only create liquidity for the market in one of two logical ways.

1. They can "warehouse" i.e., stockpile, what they are buying and selling. Under current regulations, this incurs capital charges that would make market making completely uneconomical.
2. If they do not warehouse, selling by a market maker would require a mechanism for near instantaneous borrowing of securities and purchasing. Buying would require either a large credit facility or near instantaneous financing of the bought assets.

Overall these barriers could delay "ledger nirvana" by years.

3. THE NEW APPROACH

Given the need to improve trade processing, as well as reducing the barriers to large scale adoption of DLT, is there scope for an intermediate/hybrid approach that uses some of the elements of DLT to focus on the specific causes of problems identified above?

Figure 6 presents the visual construct of what happens if you accept the assumptions implicit in the analysis above.

The model described below incorporates elements of DLT technology, data analytics, and existing software tools to meet the problems described at a technical and organizational level.

[10] http://www.bis.org/cpmi/publ/d101a.pdf

3.1 Overview

The principle elements of the hybrid solution are:

- DLT can support a mechanism that allows banks to agree on how different products will be processed. For each financial product, a "product definition agreement" (PDA) will list the agreed formats of data and the collection of services and systems that will perform the relevant parts of the trade lifecycle.
- Existing technologies would be used to load incoming messages and validate them, as well as either create a new object (in accordance with the PDA) or link to existing objects.
- It will attempt to "link" incoming notifications from the other parties to trades (or related objects) currently stored based on trade economics; it does not wait for all trade attributes to create a perfect match.
- Both sides of a linked trade are stored in the same data object (the "golden container") and any updates (except for private data) are distributed to all relevant parties.
- A "service notification" process determines whether an object has reached a sufficient degree of completeness or consensus between parties to hand it off to services (whether smart contracts, existing bank systems, or market infrastructure) that perform parts of trade processing. Conversely, if the degree of completeness or consensus is broken it will also inform the relevant services.
- The key data for analytics tools covering cost per trade, operational risk, and client efficiency will be provided by recording the capture of the trades and events, together with any exceptions, and the length of time spent processing.

3.2 PDA

The PDA provides the basic set of rules parties need to accept regarding the processing of trades they agree on, on a collaborative, distributed basis.

Figure 7: Sample product definition agreement

PRODUCT DEFINITION AGREEMENT	
Product Name	Spot FX
Product Code	FXSpot
Data Model	ISDA FXSpot
Party A Signed Date/Time	22/11/2016
Party B Signed Date/Time	19/11/2016

ATTRIBUTE OWNERSHIP	
ATTRIBUTE	ATTRIBUTE OWNERSHIP
Trade Date	Consensus
Settlement Date	Consensus
Currency Pair	Consensus
First Currency Amount	Consensus
Second Currency Amount	Consensus
First Currency Settlement Instruction	Party A
Second Currency Settlement Instruction	Party B
Valuation	3rd Party

FUNCTION OWNERSHIP						
FUNCTION	PARTY A -APP/SERVICE PROVIDER	PARTY A -APP VERSION	PARTY A -OBJECT	PARTY B -APP/SERVICE PROVIDER	PARTY B -APP VERSION	PARTY B -OBJECT
Valuation	Open Gamma	1.02	Trade	Open Gamma	1.02	Trade
UTI Exeception Resolution	Lebowski Inc	2.1	Trade	Lebowski Inc	2.1	Trade
Matching	Calypso	14.1	Trade	Calypso	14.1	Trade
Settlement Engine	Calypso	14.1	Trade	Calypso	14.1	Trade
Netting	Calypso	14.1	Settlement	Calypso	14.1	Settlement
Collateral Management	Cloud Margin	3.2	Trade	Cloud Margin	3.2	Trade
Payment Instruction	Bank Internal	4.5	Settlement	Calypso	2.3	Settlement
Operational Risk Monitor	E & Y	1	Trade	KPMG	9.7	Trade

A consensus mechanism in DLT, such as the "notary" in R3's Corda allows all relevant parties to "sign" that they are agreed to the processing rules and data structures.

The relevant parties can explicitly agree on:

- The structure of data for trades and events processed.
- The ownership of data attributes. For some data attributes, the agreement may state that one party fully owns an attribute e.g., one party may "own" the population of their own settlement instructions, which are then accepted by all relevant parties for use in the settlement process. For others, such as a valuation, a third party may have the obligation to populate the data. Finally, many attributes, such as trade economics, will be owned by both trading counterparties and the attribute is only recognized as correct when they match.
- Functional ownership records which elements of the trade lifecycle will be carried out by a specific system or party. Both parties may agree to use a specific smart contract, or a cloud version of existing vendor system, a centralized service, or they may agree to continue using their systems (accepting increased risk of differences).

3.3 The golden container

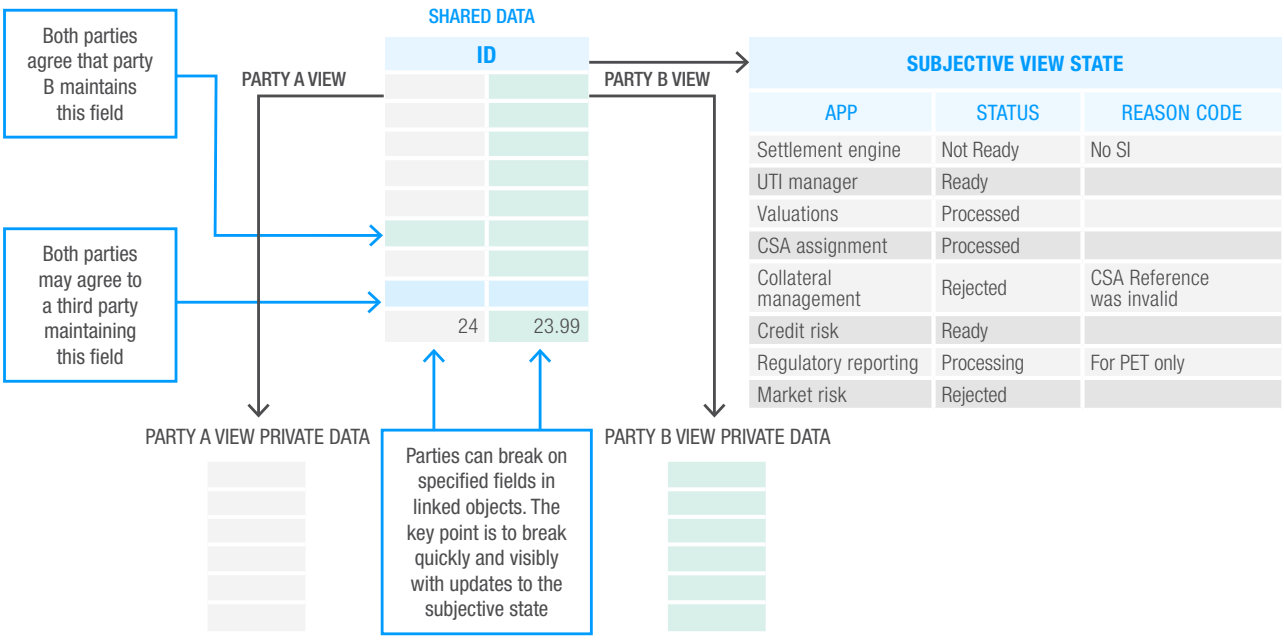
Trade and event data is stored in the "golden container." There has been a seeming endless quest by many in banks for a "golden trade" record. The general idea is to create a single version of the truth for a trade (either at an organizational or a market level). Unfortunately, the idea tends to break down unless it involves a centralized system to not only store the golden record, but also carry out all of the processing. If multiple systems carry out different elements of trade processing, including lifecycle events, enrichment, or generation of derived data (settlements, aggregated positions, etc.) there are many opportunities for the golden trade to become tarnished, even for the simplest cash products.

The concept of the golden container is very different from many DLT market's proof of concepts that assume either consistency of the trade from the outset or that the trade is either in an agreed or not agreed state. It would still use many of the strengths of DLT but would drive towards a consistent view of the trades through the rapid identification of inconsistencies between both the parties involved and their systems.

The object model attempts to balance control and flexibility. Consistency starts by linking the two parties' views of the world, providing a path to consensus, early identification of differences, and views on the state of a trade (or related object) from multiple functional perspectives.

The most important step is to create a linked version of the trade that contains both parties' views. Each time a relevant system updates a trade it updates their view of the trade and sends it to their counterparty (or any other agreed relevant party).

Figure 8: The golden container

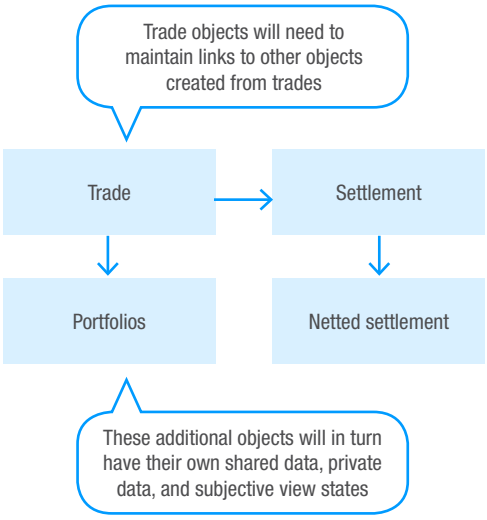


The key benefit is that both parties can see in real time if they do something to make the trades diverge.

The golden container contains two other key elements:

- **Private data** related to the trade will be stored only on the relevant party's node. Private data includes data that genuinely needs to be kept private (P&L, trading book, etc.) and data that is only needed by one party and should be kept segregated from the main trade object to avoid mutation through the addition of data superfluous to one party or the "overloading" of fields.
- **Subjective view states** make it possible to validate and process objects from the perspective of different functions. For instance, the trade may be considered valid (and ready for further processing) by one application (i.e., the front office, which requires only a calculated PV), but may not be valid from another function's perspective (i.e., the collateral management system may require the relevant master agreement to be assigned to the trade).

Figure 9: Relationship between key objects

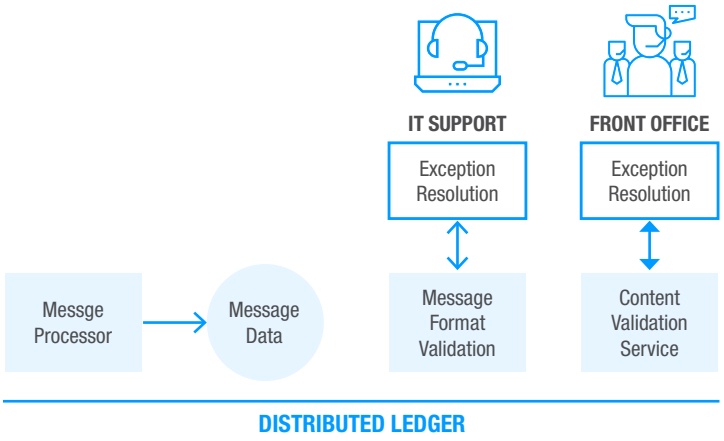


3.4 Data capture

A key element of the model is the capture of data. Data needs to be captured within and outside the bank for any relevant trade or trade event.

The first key step to adopting a more distributed approach to trade processing is to recognize the need to load trades (and other messages) into the ledger from existing sources. This could be a feed from a source where the parties are already agreed on the trade economics, such as an ECN, or from an existing trading system where trades are input manually (and unilaterally) by one party's front office.

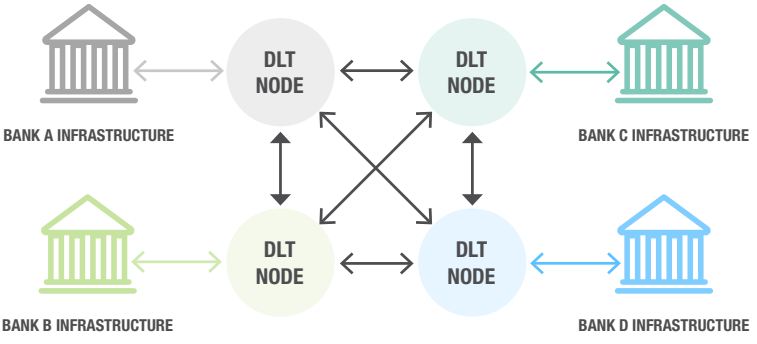
Figure 10: Data Capture



The data capture mechanism needs to be able to process any popular message format used today. To maximize flexibility, trading counterparties should be able to agree to use a third-party message validating/translation service (see PDA). There are several proven products on the market that allow the data capture and data processing tasks to be carried out, largely as a configuration task rather than a major programming exercise.

These tools would allow the processing of existing message formats used by banks, including standard formats (e.g., FIX, FpML), company specific formats, and bank specific modifications of standard formats.[11]

Figure 11: Data distribution



[11] This includes Xceptor's DataHub or Broadridge's Message Automation tools

3.5 Distribution

Nodes[12] within the firewall of each bank would control the flow of messaging between the banks' internal systems and, where relevant, to the nodes of other parties relevant to each trade. The actual secure transport layer could even be provided by an existing supplier of secure messaging that is already integrated into each bank's infrastructure.

3.6 The status monitor

The structure of this model makes it straightforward to implement a status monitor that would allow a centralized support function or middle office staff to have a near real time view on the status of a trade (and related objects) from all perspectives.

Some banks have attempted to create similar tools to benefit from potential control and operational efficiencies, but they are generally held back by the fragmented state of their architecture.

A generic trade status monitor could revolutionize the management of operational risk in the trade lifecycle and potentially allow a greater deal of standardization of process and error resolution in back- and middle-office teams.

Figure 12: Examples of status monitor

PORTFOLIO VIEW

FUNCTION	NOT PROCESSED	SENT FOR PROCESSING	PROCESSED	ERROR	NOT APPLICABLE
Trade Capture	0	234	234	0	0
Trade Validation (Technical)	0	232	231	1	0
Trade Validation (Business)	0	123	122	1	0
Matching	2	80	74	6	154
UTI Exception Resolution	31	123	123	0	200
Settlement Engine	0	143	142	1	0
Valuation	1	154	123	3	43
Trade Reporting	3	120	118	2	114
Clearing	2	32	31	1	202
Netting Engine	0	123	122	1	32
SWIFT Settlement Gateway	1	13	23	12	123

Provides a filterable view of status. This is likely to be used by operations, control functions, and IT

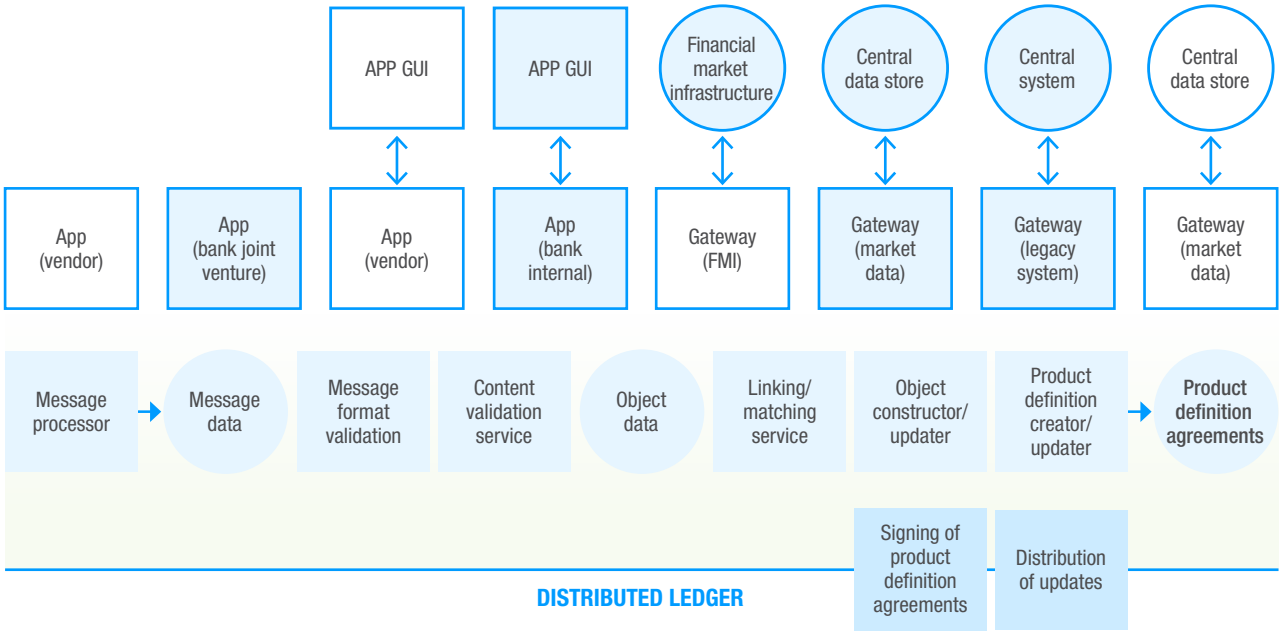
TRADE VIEW

FUNCTION	STATE	ERROR TYPE
Trade Capture	Processed	
Trade Validation (Technical)	Processed	
Trade Validation (Business)	Processed	
Matching	Processed	
UTI Exception Resolution	Not Processed	
Settlement Engine	Error	Missing Settlement Instruction
Valuation	Sent for Processing	
Trade Reporting	Processed	
Clearing	Not Applicable	
Netting Engine	Not Processed	
SWIFT Settlement Gateway	Not Processed	

The ability to drill down to trade level means anyone involved in any part of processing the trade can make a judgment on the overall operational risk. A trade level view could also display information on trade economics, etc.

[12] A system that manages the communication of messages between all parties sharing the same distributed ledger based system.

Figure 13: The overall ecosystem



3.7 Functional model

Overall, this design depends largely on the core capabilities of DLT systems and the ETL tools already in use in many banks. The new functionality consists of a set of small and relatively simple components.The "object constructor/updater" determines whether a new object, i.e., a trade or a settlement, should be created or whether an existing object should be updated.

The "linking/matching service" determines whether the new object can be linked to an object created from a counterparty's (or other third party) data. If it finds a sufficient degree of consensus it will merge the objects into a single golden container (see below) that have both parties' versions of the object in a single data structure. There will also be the option for both parties to "force match" two objects into a single golden container, if they both agree their versions represent the same trade, even if they potentially disagree about some of the details.

A "service notifier" will inform all relevant connected applications or services if the trade (or other object) has reached the state of consensus or completeness that allows another part of the trade lifecycle to take place, such as settlement or trade reporting. The service notifier will also inform the relevant services if an updated object has sufficiently changed for a cancellation or amendment process to take place.

"Product definition creator" takes new product definitions (or amendments) and manages the process of getting the agreement signed by both parties and translated into the rules that the service notifier uses to communicate with services/apps.

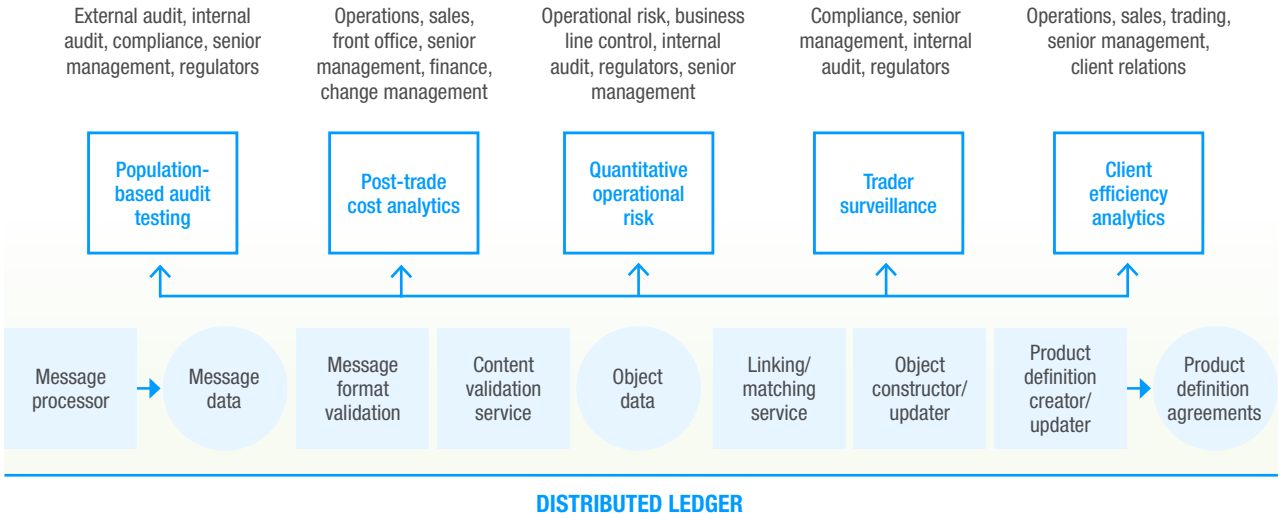
Putting these relatively straightforward components (largely based on existing technology) together creates a platform that can be introduced within existing infrastructure.

3.8 The feedback loops – cost and control

Last but not least are the "feedback loops." These will make problems more transparent, potentially cut the cost of control, and provide the basis for better management decision-making by allowing the objective measurement of system or process improvements.

Generic tools can be created using the basic data contained within the rules engine (derived from the PDA) about the "meta-workflow" of trades, such as when it was matched, when it was complete enough for settlement, and the time delays in processing

Figure 14: Integrating analytics components



The tools, which could be produced in partnerships with organizations specializing in control and/or analytics, such as audit firms, include:

Population-based audit testing: currently the external audit process is based on samples of trades and does not trust the bank's own internal records. If there is a trustworthy source of shared trade data, external auditors could test the validity of all trades recorded by one party rather than just using a sample, i.e., population based testing. Potentially, tools could be deployed that do continuous real time auditing of trade data.

Post-trade cost analytics: the main drivers of the human costs in trade processing are exceptions to straight-through processing (STP), and delays at any point in trade processing. The platform will collect the exceptions related to trades from function-specific processing apps and allow more accurate cost estimateS down to the level of the individual trade.

Quantitative operational risk: measurement of operational risk is currently a highly manual, largely qualitative process, but one that has a major impact on calculated capital charges. The platform allows a quantitative measurement based on exception rates and time delays in processing.

Trader surveillance: current systems used to detect rogue traders typically look for patterns of unusual behavior in trading and trade processing. An approach based on the use of the status monitor allows consistent business logic to be used across multiple banks to look for anomalous patterns.

> "The view of trades will be partially asymmetric between parties, because different banks have different risk appetites, different accounting treatments, or simply want to conceal information about the trade that the other party does not need to know."

Client efficiency analytics: many banks currently carry out "client efficiency" analysis to determine the relative cost of doing business with each client, using the data to change pricing or drive change. The platform provides the opportunity to collect all the relevant data from a single place, and rollout out the same analytics tool to multiple banks.

4. CONCLUSION

This paper gives an explanation of the mechanisms causing problems in trade lifecycle processing and the techniques that could be used for dealing with them. Many problems are clearly not technical but socio-economic or political in nature within banks. Nobody sets out to design bad infrastructure, but a succession of decisions, which are optimal for one specific

function or business lines, can progressively create complexity and drive entropy in the infrastructure and the organization. Complexity reduces understanding of the systems, which drives more sub-optimal decision-making, in turn creating more complexity.

One of the key factors driving this model is a recognition that in markets agreement about the "facts" of a transaction (whether between the parties involved or between systems within the same bank) can be highly unstable and asymmetric. Trade processing, in many areas of markets, is not simply about agreeing the details and settling the trade. The trade may undergo many lifecycle events and there may be changes to many of the non-economic attributes during the life of the trade. The view of trades will be partially asymmetric between parties, because different banks have different risk appetites, different accounting treatments, or simply want to conceal information about the trade that the other party "does not need to know," such as their own P&L arising from the trade. The more stable and symmetrical trade types, such as spot FX or cash equities trades, are likely to converge to the "purest" DLT model more quickly, but even they are likely to benefit from the feedback loops outlined in this paper.

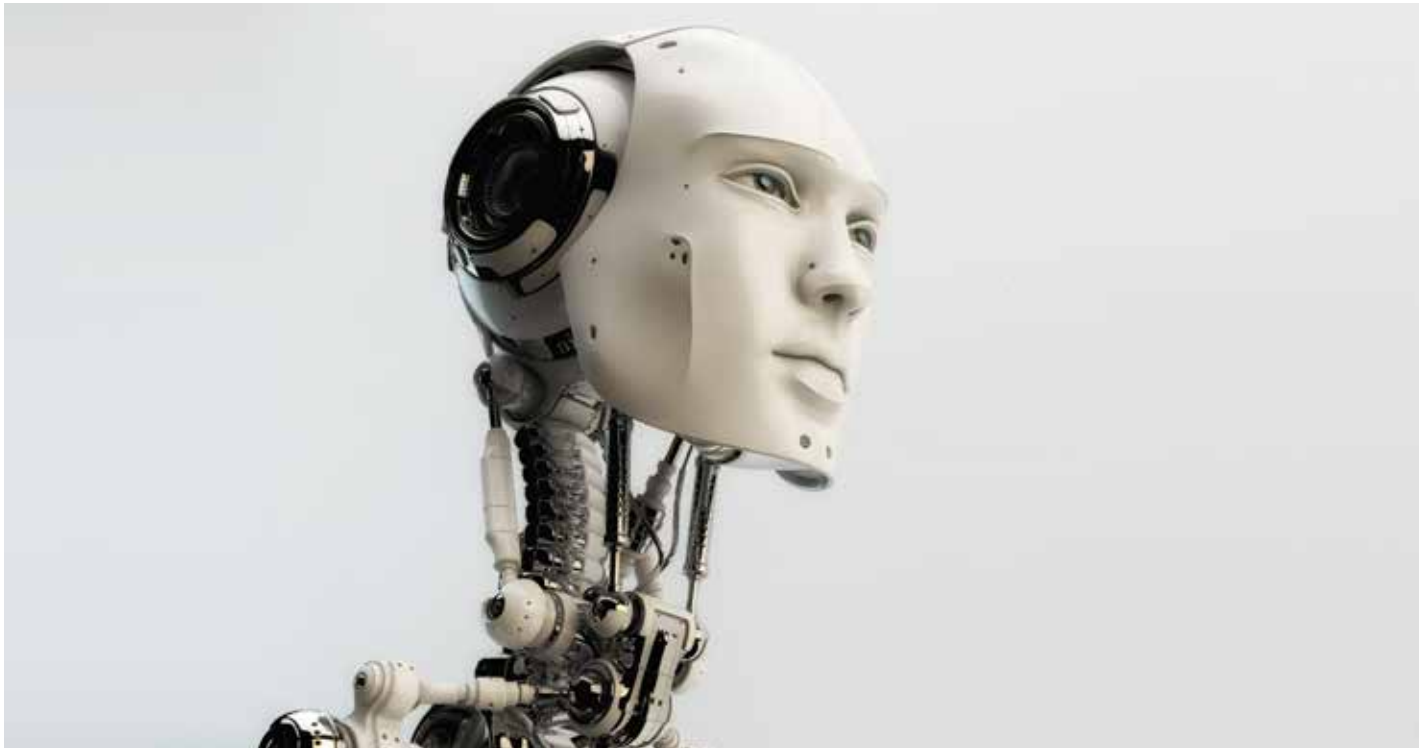
Some DLT enthusiasts may argue that there is relatively little DLT in this hybrid model, but that is by design. DLT as a technology will rise or fall based on its effectiveness in solving problems. This model introduces key benefits of DLT into the heart of trade processing, in a relatively undisruptive way. It also provides the feedback loop that can objectively measure the success of different approaches, whether they use DLT, or current technologies, or simply involve organizational or process change.

Rethinking robotics? Take a step back

ASHWIN GADRE | Partner, Capco
BEN JESSEL | Managing Principal, Capco Digital
KARAN GULATI | Principal Consultant, Capco

ABSTRACT

In these times, no conversation about disruptive technologies is complete without reference to "robotic process automation" (RPA, or robotics as it is colloquially known). Although the technology is not new – screen scrapers and automated workflow have been around for over a decade – the pace of adoption and the fact that this is now being actively experimented with and piloted in most major financial institutions is a new phenomenon. This is due to the confluence of two unique market events: maturation of robotics technology and the efforts by financial institutions to mitigate inefficiency. Similar to any rapid mainstream adoption of new technology, however, success is not always assured. In the case of robotics, adopters have faced mixed results. In this article, we examine why the adoption has been so troublesome. We will explain why the institutions that have been successful in adopting robotics have done so not by focusing on the technology but by taking a step back and looking at the actual business problem at hand, and then considering robotics as part of a broad toolset that is available to them.



1. INTRODUCTION

Since the financial crisis, financial services firms have been engaged in a war against inefficiency. With a low interest rate environment and increased regulation, margins have become compressed to the point that the biggest driver to profitability is now a financial institution's ability to control its costs and improve efficiency.

In retail banking, a low interest rate environment along with caps on transaction fees mandated through the Durbin agreement have severely limited the sector's ability to generate revenue, whereas the increasing cost of regulation (estimated at over U.S.$70 billion[1] since 2011), inefficient processes, and aging technology have been a drag on the bottom line.

As a result, banks' costs to maintain an average checking account are nearly U.S.$349, yet on average they generate only U.S.$268 in revenue for each account. This leaves banks needing to find nearly U.S.$81 per customer in either increased revenue or cost efficiencies to make up the difference.[2] A similar story exists in the corporate and investment banking world – the revenue drivers of pre-2008 are no longer there, yet the cost burdens have continued to increase.

Banks had responded to margin erosion before, in the 1990s and early 2000s, by moving their non-value added operations offshore to India, the Philippines, and other low cost locations. Consequently, many banks have already optimized their operations from the perspective of minimizing the cost through labor arbitrage.

However, having a cheaper workforce is one thing, doing more with fewer people is another. Financial institutions are now beginning to realize that having outsourced many of their back- and middle-office capabilities to other organizations that run them in different locations means that they have lost control of their ability to optimize the process themselves. As a result, we are seeing financial institutions starting to take control of their operations and look for ways to both drive efficiency out of their processes and replace humans with technology through automation.

The opportunity for financial institutions is clear by looking at the divergence between highly efficient banks and those that have yet to grasp the nettle of becoming

[1] http://on.wsj.com/2xcmYIK
[2] http://bit.ly/2fGCjv3

more efficient. Benchmarks clearly demonstrate that the difference can be as much as a four-fold increase in efficiency with leading institutions able to serve over twelve hundred customer accounts for each employee in comparison to ones at the bottom of the efficiency heap who are only able to support a quarter of that amount.[3]

2. ENTER ROBOTICS – POTENTIAL TO UNLOCK OR AN OVERHYPED TECHNOLOGY?

With so much pressure on finding a way to drive efficiency, financial services firms are looking at new technologies, such as robotic process automation (RPA) to help them bring down their costs. While robotics is not new, it is one that has evolved to a point that financial institutions see it as a technology that has now reached maturity and has – for the most part – been proven.

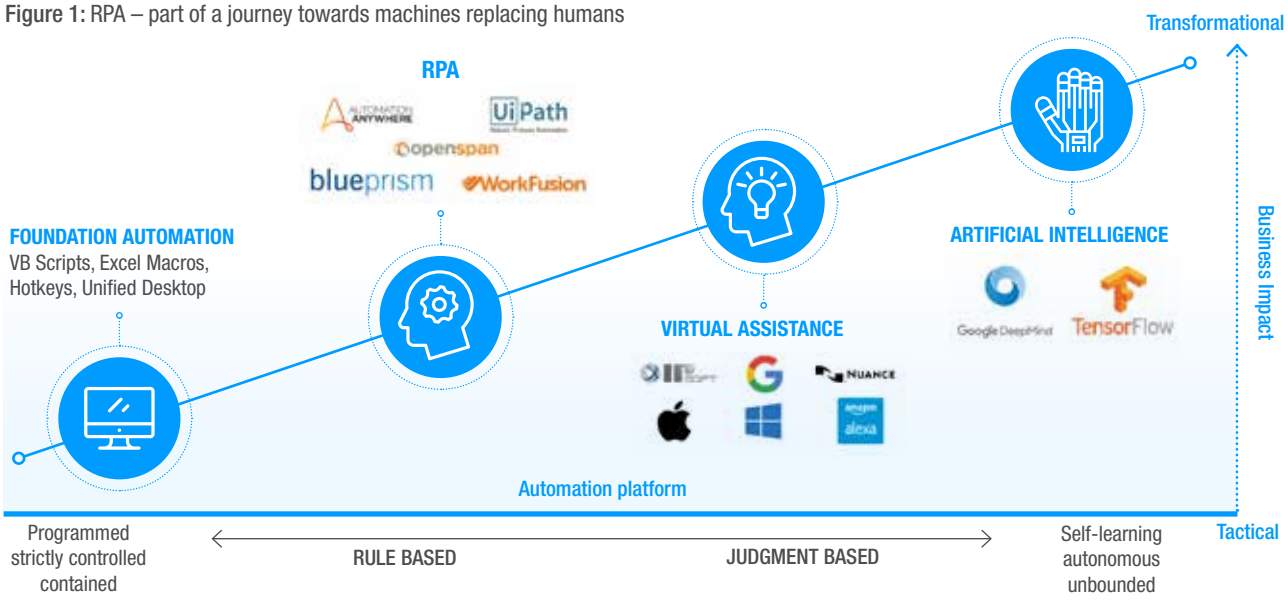
The technology itself started from humble beginnings through small scripts that were written to repeat certain jobs or to provide quick integrations between systems that did not have any interoperability. Over time, the technology has evolved to one that is "fourth generation," which means that increasingly tasks that used to take many lines of computer program could be carried out by clicking and dragging icons in a software program, with the underlying code and configuration being developed by the software package.

RPA tends to be used to integrate systems where it is too costly to integrate them at a deeper level and so the technology is often cited as a tactical band-aid. While that may be the case, it has a significant role to play in enabling organizations to achieve operational efficiency.

Because RPA requires rules to make decisions, it is ill-suited to the kind of human decisions that require an element of judgment or where a business problem involves making sense out of large volumes of data sets, deriving a rule based on examining correlations. That technology is within the realm of cognitive agents, machine learning and artificial intelligence, which is an immature field that is rapidly evolving.

However, mainstream adoption is in its infancy and the jury is out on its efficacy. As is usual with this part of the hype cycle, there is an equal weight of claim and counterclaim concerning its ability to drive out costs, with Gartner reported as claiming that 96%[4] of organizations achieved benefits from robotics whereas other organizations have reported that less than half of robotic endeavors have succeeded. Our own experience has been that more appear to be failing than succeeding.

Figure 1: RPA – part of a journey towards machines replacing humans



[3] http://on.bcg.com/2hLorE8
[4] http://bit.ly/2xdTOOe

Notwithstanding the lack of a single empirical view on the success of the technology, we have formed a viewpoint, from our experience of being in the market, as to what separates the organizations that successfully adopt robotics – whether they be the 4% or the 50% – from the failures.

3. SEPARATING THE BOYS FROM THE MEN – WHAT SUCCESSFUL ORGANIZATIONS DO TO SUCCEED IN PROCESS AUTOMATION

We have surveyed numerous organizations and identified common traits among firms that have been successful in improving and automating their operations.

3.1 They focus on the process, not the technology

The idiom of "when your tool is a hammer, everything looks like a nail," rings particularly true with robotics. And what we mean by that is that organizations that struggle with robotic adoption are the ones that approach it from the perspective of the technology, not from the business problem. The problem with this is that these organizations end up focusing on the wrong problems to solve, or always attempt to solve everything with robotics as opposed to being discriminatory and understanding that robotics is just one tool in a toolkit of multiple approaches to taking out costs and removing inefficiency.

One of the most common scenarios where we see this is where an ambitious technology executive looking to demonstrate to the business that they are looking to provide the business with value will procure a robotics tool (more often or not Blueprism or Automation Anywhere). They will issue a mandate to the technology organization to identify opportunities to use it. It is certainly a noble ambition, but these initiatives tend to wither on the vine as the technology team is not able to understand the business domain sufficiently to identify what problem to solve, and are unable to get buy-in from the business, who are usually confused as to why the technology team appears to be forcing the tools on them.

This can also happen from the business side of the organization, and while results can be better, going hunting for a problem armed with a solution rarely leads to a satisfactory outcome.

We commonly see this with the advent or robotics labs within organizations, which are designed as "Centers of Excellence" for the use of robotics in the organization. While there is nothing wrong with the establishment of such an organization per se, the danger comes from how the lab then engages the business community – they often do it from a technology solution perspective rather than from a business challenge one.

This challenge has become exacerbated by the proliferation of cheap robotics tools that are easily available. This is because a potentially dangerous technology is placed into the hands of people in the organization that may not have the sophistication or structured thought process to fully appreciate that the best way of solving a business challenge is through deconstructing the business problem as opposed to starting with the technology, which leads us onto the second aspect that leading organizations do well.

3.2 They understand the business problem they are trying to solve

Frequently, seemingly straightforward processes are in fact far more complex than they look at first view, and many organizations in their enthusiasm to adopt robotics do not give enough time to fully appreciate some important nuances. This can be because they do not get business SMEs involved or are just too aggressive in their approach.

This leads to many problems: The wrong processes end up being automated or the right processes are not automated properly. One of the most common issues, here, is that the robotics team will solve one bottleneck in an overall business process, only for the workload to flow to another bottleneck with the result that the organization ends up spending significant amount of time and money to end up with little improvement in throughput.

In other cases, robotics teams misunderstand the level of implicit decision-making and human judgment that a process actually needs, ending up with a process that takes far longer than it did before it was automated.

This is because robotics is a rule-based system and robots cannot make determinations based on their own judgment. When a robot identifies an input for which it is not equipped with a rule, it has to escalate a case to a human in an exception queue, where the issue must sit until it is resolved. An exception queue is designed to provide a human level of support in processing when

robots face challenges. These tend to be staffed by generalists, since many of the specialists that the robot has been designed to replace have left the organization as their roles have been automated. When poor robotic design ends up with generalists having to spend more time than expected, the exception queue rapidly becomes a bottleneck in the process. Ultimately, this leads to a slower process than before automation was even considered.

One way that successful organizations can mitigate this is by bringing the business process into a lab, where staff that have the responsibility of improving and automating a process can work side by side with an operations professional. One of the benefits of such labs is that the optimization team get to have a ring-side seat on how a process is carried out versus how they are told it should be carried out by a process owner who graduated from a hands-on operational role some time ago and may no longer be in touch with the actual realities on the ground.

3.3 They involve technology stakeholders in their automation efforts so that they avoid the dangers of "shadow IT"

The process of making a technology reliable and supportable is a slow, and complex one, and is often misunderstood by the technology department's business stakeholders who just perceive technology as being slow and unresponsive.

With robotic platforms that can be deployed and built upon without the involvement of the technology department, a lot of power is placed in the hands of the business, enabling them to circumvent the technology department. While the business may delight in the ability to rapidly deploy robotics, without relying on technology stakeholders there ends up being severe ramifications down the road because of supportability of the robotics solution.

Not only does this technology become invisible to the technology team, whose mandate is to ensure that technology that supports business process is reliable and managed, but it also introduces a fragile technology into the organization. Robotics is fragile, as it integrates the user interfaces of "line of business" and corporate applications. One small change in one of those systems – such as changing the name of a field on an application form – can break a robot!

In some cases, these robotic workarounds developed by non-technology teams end up supporting business

critical processes. When they break, organizations have the twin challenge of having to fix a critical business process and having a technology team that has no knowledge of the robotic system, and, therefore, are unable to fix it. With regulatory requirements placing increasing requirements on organizations to fully document the presence of business critical technology, the rise of this technology is a major concern for the technology and compliance departments.

Successful organizations understand that technology needs to be part of the solution from design through to implementation and support. Getting buy-in from technology is hard because they do tend to view the technology as inherently unstable. In addition, in many cases it is a way of pushing investment decisions concerning replacing or integrating technology down the road, which can oftentimes run contrary to the technology department's desire to avoid quick-fixes and move the organization off legacy technology that is expensive to support.

3.4 They are in it for the long haul

Many organizations make the mistake of assuming that they will be able to receive a return on investment (RoI) that will be comparable to other projects in a change portfolio that are built with mature technology. Similar to the adoption of lean, six sigma, and kaiban, which required a mandate from the top, and many years of integrating the discipline into the organization at all levels, robotics will need to take the same path. We frequently experience conversations with our clients where they state that they have struggled to get buy-in for their robotics program because they cannot justify the RoI.

In the same way that to climb mount Everest you must start with the foothills, successful organizations understand that the adoption of robotics is a long journey that starts off with simple projects with limited upside that builds towards a sophisticated efficiency generating capability over time.

3.5 They actively manage and do not "set and forget"

For organizations that have managed to get as far as implementing a robot in a production environment that is supported by IT, there remains one final hurdle that concerns the misbelief that a robotics journey ends with the transfer into production support. In fact, robots need almost as much care, attention, and management as the real humans that they replace, or now sit aside.

Robotics is not a panacea – small changes to line of business systems can render these robots inefficient, and in some cases, cause them to break. Furthermore, they are often not effective in circumstances where there is a level of ambiguity. For example, it is common for a robot to be unable to distinguish between a $ sign and a five.

Whereas a human being can make the distinction through either better visual recognition or by looking at the context of the overall document or record that it is situated into and make an educated guess, a robot will often need to refer this to an "exception queue" of humans for further review.

The implication here is that robotics, in some cases, can actually be less efficient than humans and cause bottlenecks. Robotics is an inexact science, and the nature of the information that the tools have to deal with is often variable. Consequently, the discipline of monitoring the effectiveness of a robot or team of robots needs to be effectively built into any robotic operation. One has to recognize that robots will fail, they will need care, and attention and without this management, they will cause more harm than good.

3.6 They leverage their existing technology

Robotics is an effective tool for integrating lines of business and corporate systems, but it is not well suited as a business application in-of-itself. It is a form of band-aid where the cost of integrating systems that are silo-ed, and are not interoperable, is too high. That is why robotics must be approached from the perspective of leveraging existing technology with robotics providing the glue.

We have seen organizations mistakenly attempt to replace core applications with systems they have built through robotics platforms. Robotics is not an effective long-term tool for this kind of approach.

For example, some organizations attempt to replace their operational ticketing system with a home-grown robotic solution. In most cases, it is far better to leverage an existing system's functionality, and integrate the robot into the application than to build a new system from the ground up using robotics.

4. FOCUSING ON THE PROCESS FIRST, NOT THE AUTOMATION

With automation being so in vogue, we have seen a tendency in the industry to focus on automation as the solution for resolving business problems and achieving efficiency. Instead of taking a methodical approach to focus on the business challenge, many are looking at ways of improving the process and then looking at automation as part of a broad array of tools that can be applied to generate business value.

Oftentimes, simply reviewing and improving a manual process can be far more effective than automating it, for reasons such as the need to use human judgment, ensure a degree of customer intimacy, or for quality control or regulatory reasons.

Optimizing a process should always be the first port of call, irrespective of whether the process is eventually automated or outsourced. The famous outsourcing idiom "your mess for less" was quip that was often used in the 1990s to describe the pitfalls of outsourcing or automating a process without fixing the underlying problems associated with the process first. The same applies to automation: automating a bad process just increases the speed of failure and inefficiency. In many cases, we have seen process times increase as a result of organizations attempting to automate a bad process, or one that transpires as being inappropriate for robotics.

Many thousands of column-inches have been written about business process automation over the years and while we will not cover them here in detail, some key aspects are important to recap as they are foundational to an overall automation journey.

Business process optimization concerns the identification of a business issue, and recognition that it lies within a domain of problems that involve processes, skills, data quality, and governance. It requires practitioners to holistically assess a function and sub-processes, to identify the best strategy for implementation. Process optimization is about taking a top down view of all key business units in a functional area and defining the landscape in which they operate. It should help answer the following questions: is the process efficient? Are there bottlenecks? Do all parties have quality data that are accurate and actionable? Are there vendor dependencies that create risk for clients? Are there transformation efforts currently underway in a business area?

It is important to answer these questions because frequently organizations seek to fix – and in some cases, automate – the wrong functions. There is ultimately a balancing act between the effort required to improve a process and the value that it will bring to the organization. It is not just the time taken to rewrite operational manuals and retrain staff; there are also changes to technology, organizational resistance considerations, as well as risks associated with optimizing business critical processes.

Ultimately, there is a prioritization activity that process optimizers go through to determine whether the benefit is worth the cost. Usually, this filtering results in an extended list of process areas to automate, which requires further definition through documenting a detailed current state process map that captures complexity and value add time for each step, key interactions between teams, dependencies, and automation potential.

Figure 2: Taking a structured approach to process optimization

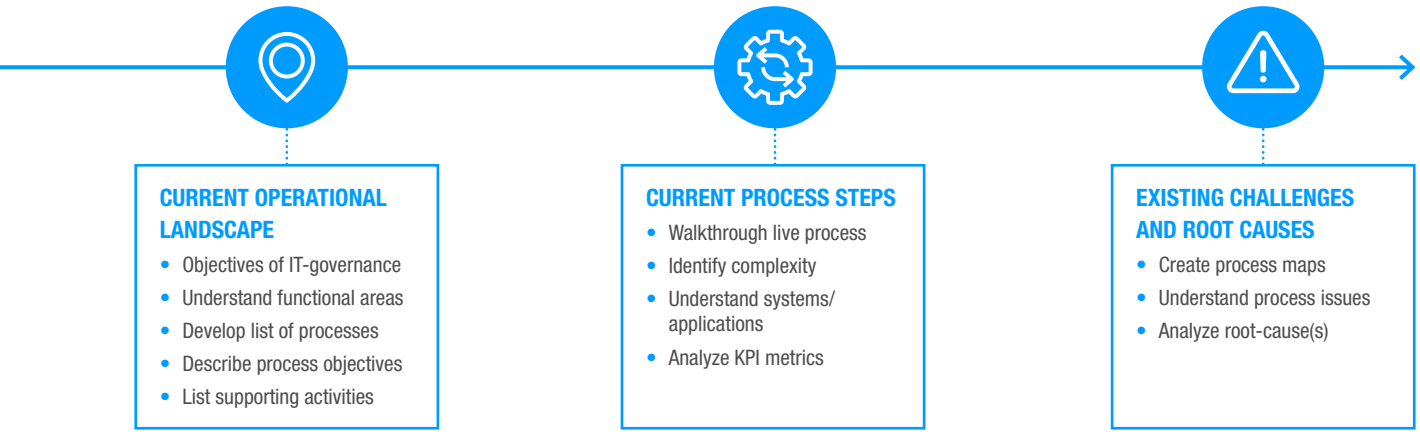
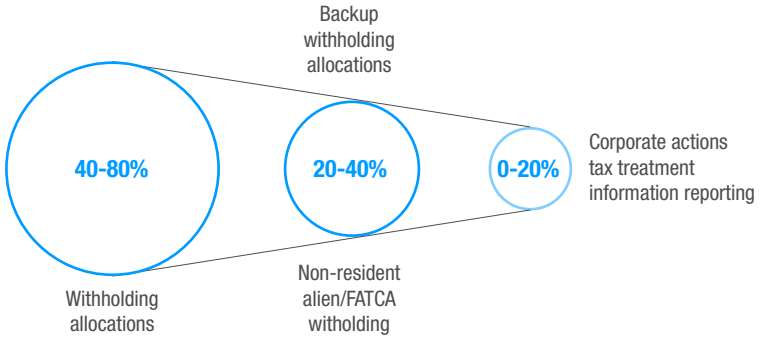


Figure 3: Typical savings of automation applied to various aspects of tax operations



Creating this artifact at a granular level is a "value stream exercise" that analyzes a function from the top down. The value stream "map" incorporates additional information, such as a system landscape that highlights the number of unique platforms and applications within a process. This level of detail helps to identify dependencies and pain points that originate from sources not previously considered, such as third party applications.

Often this can result in surprising discoveries as to what the root cause of an issue is, which could be operating model issues, system or infrastructure gaps, data quality, and workforce management, to name a few.

Taking a holistic view to process identification allows an organization to drive meaningful conversations on automation opportunities, and strategize on a possible suite of optimized solutions. Considering linkages and dependencies between functions, processes, sub-activities, and systems increases your odds of identifying and addressing the right problems.

5. IDENTIFYING THE RIGHT PROCESSES TO IMPROVE AND AUTOMATE

Broadly speaking, automation candidates fall into several key areas, such as document heavy processing, processes that involve comparisons and data entry between different systems, especially in time critical situations, and customer service functions that deal with high volumes of very similar inquiries.

5.1 Automating document heavy processes

Document heavy processes, where data needs to be extracted from structured documents and then entered into systems, are strong candidates for automation, provided that the documents involved are structured and similar.

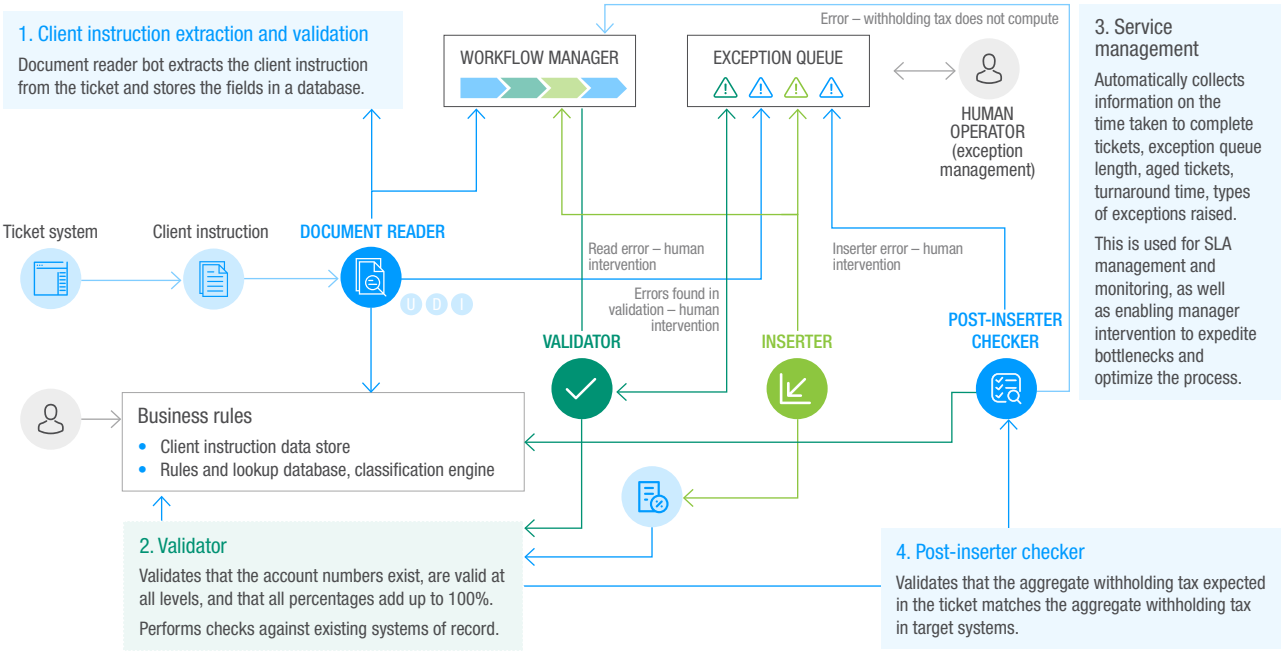
Typically, processes that involve humans extracting from documents are not only inefficient, but introduce significant risk into an organization because human error is common, especially where they have to perform tedious repetitive tasks.

This type of process is often found in tax operations functions in banks that are managing or administering funds on behalf of their clients. A key aspect of this management involves ensuring that the tax treatment of their clients is managed effectively. In this scenario, clients provide instructions on a regular basis to their financial institution that articulates how funds should be allocated for tax treatment purposes. The institution calculates the tax to be levied for a given fund, including how much tax to withhold – in the case of non-U.S. beneficiaries – as well as determine the tax treatment for corporate actions associated with the underlying securities in a client's portfolio.

Accuracy is extremely important, as any errors can have a significant effect on the how much tax a client must pay and reclaiming of overpayment of tax from the IRS is a complicated and time consuming matter.

Typically, these financial institutions seldom stipulate to their clients as to what format to use, which means that there is significant time spent by operations staff in deciphering instructions, and working out what information to extract. Furthermore, the work can be very repetitive – we have seen clients submit instructions that contain anything from one to 10,000 instructions, which can take anything from a day to a

Figure 4: Example robotics architecture for tax operations



month to process purely based on the time that it takes to copy information from a document and paste it into a form of a tax platform.

Automation is a natural candidate for some aspects of this; it is perfectly feasible for a robot to extract and input data into line of business systems in seconds, as opposed to months.

However, this is only possible if the documents that the robot is working from are identical, in terms of structure, which in the real world is seldom the case. Robots can be trained to recognize documents and work out where to extract the information from, but they need to be trained on each type of document, which can be a significant overhead.

Usually, these situations follow an 80/20 rule – the majority of documents are indeed similar, with the least amount of volume involving documents that tend to be different. The key to successfully automating tax operations resides in the ability to understand where the efficiencies lie and focusing efforts on where the data is standardized, leaving areas where it is not for humans to handle.

Not all areas of tax operations are good opportunities for automation, such as in corporate actions processing, where financial institutions need to determine the tax treatment for a given security that has undergone a corporate action, such as a stock split, dividend payment, or takeover.

Corporate actions have forever been the bane of financial institutions because there is often inconsistency between data feeds that report corporate actions, and the tax treatment of securities involved requires a high degree of objective human judgement that draws upon prior experience and reasoning. While robotic systems perform well in environments where they can follow clear robust rules, they struggle in situations where the decision process is more complex and less rules driven, as they tend to be in this case.

As we have seen, while tax operations provide some significant opportunities for automation, there are many issues that need to be taken into consideration, such as the degree to which decisions require judgment and experience, as well as the extent to which instructions from clients tend to be sufficiently standardized in their format.

Tax operations tend to be good candidates for automation, as they are large functions that can involve hundreds of staff. Hence, a small uplift in efficiency can build a business case relatively easy. In our experience,

as a rule of thumb, automation yields benefits when the amount of staff in scope for automation is within the fifty-to-one-hundred range. Departments with fewer tax operations specialists tend to not yield a sufficient saving to justify the cost of implementing automation.

5.2 Data entry between multiple systems

Currently, business processes in financial instructions tend to be supported by a myriad of line of business systems that are often not integrated. They exist as siloes of information, with humans providing the heavy lifting of ensuring that the data in one system is correctly copied into another.

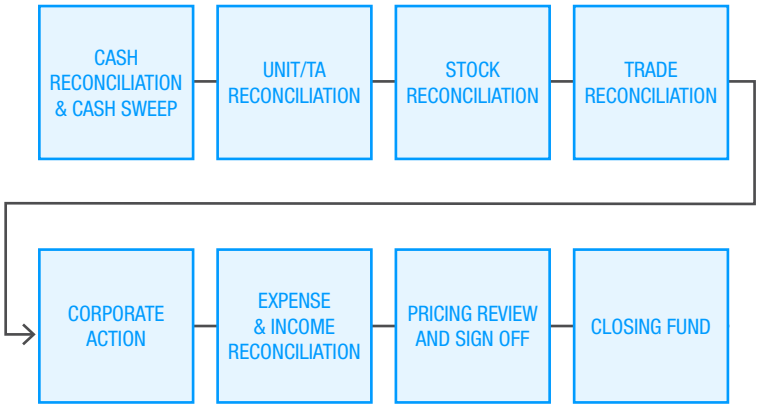
This is a particularly common phenomenon in fund accounting. While in theory, modern accounting technology should have rendered the role of fund accounting to obsolescence, the reality in most major fund management organizations is that the fund accountant's role today is one of checking that the various siloed and independent systems that report on the various part of the business correctly reconcile.

Furthermore, not only do fund accountants need to be content with consulting and reconciling data across multiple systems, but they must do so for a process that is time critical; fund accountants typically have no more than a few hours from receiving pricing information once markets have closed to "striking a NAV," whereby they calculate the value of the funds that they administer.

Even in technology progressive fund management organizations, the process is laborious and fraught with challenges concerning data quality, and the timing of information provided by other teams they are dependent on. As with tax operations, corporate actions processing plays a significant role in the calculation process, and the interpretation of the treatment of associated securities can also be subject to human interpretation.

Whereas the aspect of the process involving checking that siloed systems reconcile is a natural fit for robotics, the determination of the underlying reasons for data not reconciling and resolving this tends to be a challenge that is less suitable for robotics, as there can be so many different underlying causes for the records being different. There are currently many tools on the market that already perform sophisticated reconciliations, and the danger of attempting to solve for reconciliation in fund accounting is inadvertently investing large amounts of money to develop a solution where there are better options available in the market.

Figure 5: A typical fund accounting process



The key to success when automating fund accounting, therefore, becomes the art of being able to not just automate the comparison of data between systems, but to also address the common aspects of reconciliation associated with the root causes of data issues between systems, while leaving the more complex aspects to human beings or packaged solutions such as FIS' IntellimatchTM to resolve.

An additional factor is automating processes that lie within the critical path of a process. With fund accounting, much of the work effort occurs within a compressed timeline towards the end of the day. It is these activities that dictate the overall staffing required to support the fund accounting function, and the effort put towards automation of activities outside this window are often wasted.

5.3 Contact centers/customer service

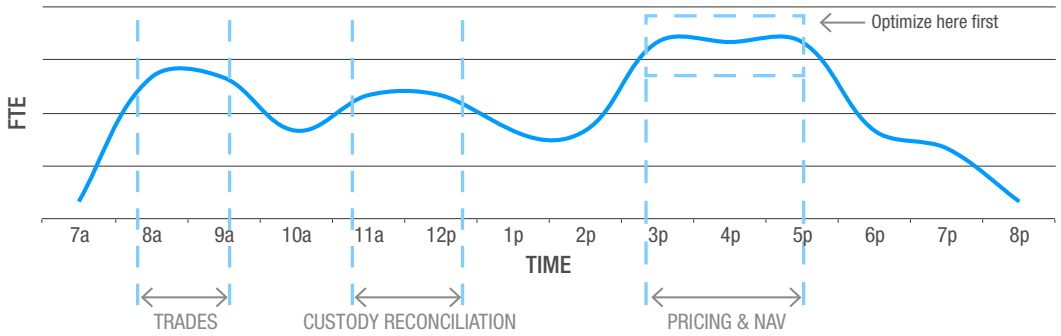
Customer services is a significant cost for financial institutions, especially those that have a large customer base, such as in retail banking or wealth management.

Customer service and contact centers have been targets of automation for nearly a decade. Almost every phone interaction with an organization, financial or otherwise, tends to involve "interactive voice response" (IVR) systems that triages call, much to the annoyance of the caller. It is a minor irritant to the caller but a significant cost saving for the bank and generally considered as the price worth paying for the cost of free banking.

However, the stakes are far higher in wealth management, where it is entirely possible to, on two consecutive calls, handle a client with a net worth of

Figure 6: Automating the right part of the fund accounting process

Typical operations of a fund accounting organization – why the NAV window is the focus of optimization



Ultimately the pricing and NAV window tends to have the highest workload and time criticality, and drives the overall staffing for fund accounting. This is why we focus on process improvement and automation initially in this window before looking at the next highest area

$10,000 and another with over $10 million assets under management.

Clearly, financial organizations want to prioritize the clients that generate them the most revenue. It is also important to prioritize inquiries that are administrative by nature – such as inquiries about balances – and those where the bank has an opportunity to increase the revenue they receive from their client.

Separating these calls is not easy; IVR is a blunt tool to use to stratify customers and types of inquiries and has the potential to annoy and deter wealthy clients.

We have seen wealth management firms setting lofty goals to drive all administrative calls to a virtual robotic agent, with the remainder handled by a specialist.

In addition to this, some have gone a step further and used automation to provide customer service agents with data about the customer, which enables them to provide the customer with a personalized service. These customer-relationship-systems-on-steroids provide an instantaneous view of profitability for a given customer, a view on open cases that the client has raised, as well as recommendations of products and services to suggest to the client based on their specific situation, risk appetite, and investment profile.

This is the true frontier of customer services – focusing on the calls that generate value for the customer and the institution and then providing the client with an experience that makes them feel like the organization knows them personally.

The use of IVR technology to help clients with administrative questions is increasingly being replaced with intelligent virtual assistants that have cognitive capabilities that go beyond simply reading from a script and triaging based on a pre-defined set of responses. Technologies, such as IPSoft's Amelia, are capable of being trained in much the same way as human can be. They can read and learn operating procedures, take large documents and synthesize their meaning, and provide answers to questions that they have not been explicitly trained to ask, purely through their cognitive "human like" capacity to learn through reading.

With customer services, we are starting to approach what futurologists have coined "the uncanny valley" – a point where computers resemble human beings so closely that they are nearly indistinguishable from each other. However, the technology is far from perfect, and organizations that have been successful in adopting this technology have been able to delineate the mature aspects of the technology from the cutting-edge elements.

6. CONCLUSION

Robotics is at the peak of a hype cycle where claims of its efficacy and ability to transform organizations are characteristically overblown. We are seeing first adopters approaching this technology with a degree of inconsistency in their success, which has been due to a combination of unrealistic expectations and taking a myopically technological approach to solving problems, even when the solution is far better suited to a non-technology route.

We have also seen how leading organizations have approached automation in a balanced way, approaching it from a holistic process perspective, based on understanding the problem and assessing the overall business case associated with fixing an issue.

There is clearly a very large amount of inefficiency in the business process of financial services across the front-, middle-, and backoffice, which can be significantly reduced through a combination of process automation, improvements to the overall operating model, and improving the consistency of data.

In many cases, there will be scope to take the additional step of introducing automation. But that will require more than a business case to drive forward. Only once the organization has articulated an intent from the senior ranks of the company to embark on an automation journey – that also takes a structured approach to evaluation that is inclusive of both the business operations and technology stakeholder groups – should such a decision be taken.

But at that point, the step forward is not merely an incremental one, it is – in the words of Neil Armstrong – a giant leap, which will enable these enlightened organizations to significantly differentiate themselves from the competition.

To robo or not to robo: The rise of automated financial advice

THOMAS H. DAVENPORT | President's Distinguished Professor of IT and Management
Babson College, Research Director, International Institute for Analytics, and Digital Fellow,
MIT Center for Digital Business

ABSTRACT

Many financial firms are now providing online and somewhat automated investing and financial advice to their customers. These "robo-advisors," some of which include roles for human advisors, are growing rapidly and generally charge lower fees than human advisors. The history, key functions and processes, and likely future directions for automated financial advice are described. Implications for human financial advisors are also discussed.

1. INTRODUCTION

The provision of financial advice to consumers is increasingly being automated. Instead of a conversation with a financial planner, investment advisor, or broker, many consumers are increasingly receiving digitally-based recommendations that are personalized to the individual. In many cases, the recommendations are implemented automatically. These investment recommendations, financial and wealth management plans, and operational financial alerts are now commonly dispensed to middle-net-worth individuals and families, from the millennial generation to baby boomers.

What does this approach mean over the longer term for the financial services industry? What changes may take place as machines grow increasingly more intelligent, and as increasing amounts of online data about personal finance becomes available? What does it all portend for human financial advisors? I will address these and other issues about automated financial advice later in this article.

These systems are often referred to as "robo-advisors," although the term is often reviled within financial firms. This is sometimes because the firms are employing hybrid human/machine solutions (discussed below), or perhaps the term "robotic" suggests overly structured and simplistic advice. In any case, I will refer to the field as "automated financial advice," even though in many cases it is only partially automated.

Many firms have adopted some form of this digital, automated, or semi-automated advice for investing or wealth management. Startups like Personal Capital, Betterment, and Wealthfront offer primarily online offerings. "Self-directed" investing firms like Vanguard, Fidelity, and Charles Schwab have had them in place for several years. Brokers including Morgan Stanley and Merrill Lynch have recently announced an advisor-mediated system. Traditional banks like JP Morgan Chase, Wells Fargo, Bank of Montreal, and HSBC have rolled out or announced robo-advisors. And even high-end wealth managers like UBS and Goldman Sachs have some form of automated offering.

However one refers to the concept, automated financial advice is growing rapidly. A Deloitte study [Srinivas and Gordia (2015)] estimates that assets under automated management in the U.S. may grow to U.S.$5 trillion to U.S.$7 trillion by the year 2025 from about U.S.$300 billion today. This would represent between 10% and 15% of retail financial assets under management. At the end of 2016, Fitch Ratings estimated that all robo-advisors managed under U.S.$100B in assets, and predicts double-digit growth in assets under management over the next several years [Reuters (2017)]. One consulting firm, A.T. Kearney, predicted that assets under "robo-management" would total U.S.$2.2 trillion by 2021 [Epperson et al (2015)]. The prediction was based on a study of consumers, many of whom expressed interest in automated financial advice.

These predictions suggest that while traditional human advice isn't going away, any firm interested in wealth management cannot afford to ignore automated advice.

> "Assets under automated management in the U.S. may grow to U.S.$5 trillion to U.S.$7 trillion by the year 2025 from about U.S.$300 billion today. This would represent between 10% and 15% of retail financial assets under management."
>
> [Srinivas and Gordia (2015)]

2. THE CONTEXT FOR FINANCIAL ADVICE

A number of trends have converged to make automated investing advice possible. Demographic trends in many wealthy nations suggest aging populations with increasing longevity, which creates anxiety about outliving resources in retirement. In the U.S. and several other countries, the move away from corporate pensions means that employees are responsible for their own investment decisions. As investment options become more numerous and complex, individual investors need more help in making decisions, but many cannot afford to pay a human advisor.

In the investment landscape, an important trend favoring automated advice is the move to passive investing. When clients invest in index funds and ETFs, it is much easier to construct portfolio recommendations. Since 2010, money has flowed strongly into passive investments more than active; in most years active flows were negative or flat. In addition, the majority of active firms have lagged behind their chosen benchmarks in investment performance over the last decade [Ellis (2017)].

Another key trend favoring automated advice is that information about financial markets and products

has exploded. Much of it is available for free or at significantly lower prices than Bloomberg, for example, which charges for a terminal [Weil (2017)]. This makes it both more difficult for any investor to gain an edge in price discovery, and makes the use of computers and algorithms more important to digest all the information.

The great majority – more than 90% – of active trading in stock markets is by institutions and professionals. Individual investors have a number of disadvantages in competition with them, one of which is the extensive use of analytics and algorithms as the basis for trading. Even the most sophisticated asset management in hedge funds is increasingly driven by algorithms. Hedge funds that use algorithms for trading already account for almost a third of the industry's assets, according to Hedge Fund Research, Inc., and quantitative hedge funds have outperformed other types over the last 25 years [Mackintosh (2017)]. Since amateur investors are unlikely to be able to compete effectively with such analytical prowess, they are probably more likely to turn their money over to professional advisors (machine or human-based) or investment firms.

Regulatory factors are also helping to drive automated advice. Fiduciary requirements for retirement-oriented financial advisors are now in place in the U.S., which may lead investment and wealth management firms to put algorithms and automated rules in place to ensure advice in the client's best interest [Fuller and Patrie (2017)]. While there is some doubt that a fully automated system can be classified as a fiduciary, most observers believe that a hybrid human/machine advisor can be [Tergensen (2017)]. In the U.K., automated advice has been given impetus as a means to provide low-cost and customized investment advice by the Financial Conduct Authority's Retail Distribution Review [Europe Economics (2014)]. A review of financial and investment regulation by asset management firm BlackRock suggests that most jurisdictions that have commented on automated or digital advice have been positive or neutral on the concept [Novick et al. (2016)].

3. TECHNOLOGICAL PRECURSORS OF AUTOMATED ADVICE

Several precursor technological components of automated financial advice have been developed over the past couple of decades. William Sharpe, a Nobel Prize winner in economics, developed the first automated financial advisor in 1996 and co-founded the firm Financial Engines [ThinkAdvisor (2015)]. The company primarily serves workplace retirement programs and employs Monte Carlo simulation to calculate the probability that an investment portfolio will meet financial objectives given many different market outcomes over the next 30 years. Several other firms have adopted the Sharpe simulation approach in their own automated advice systems.

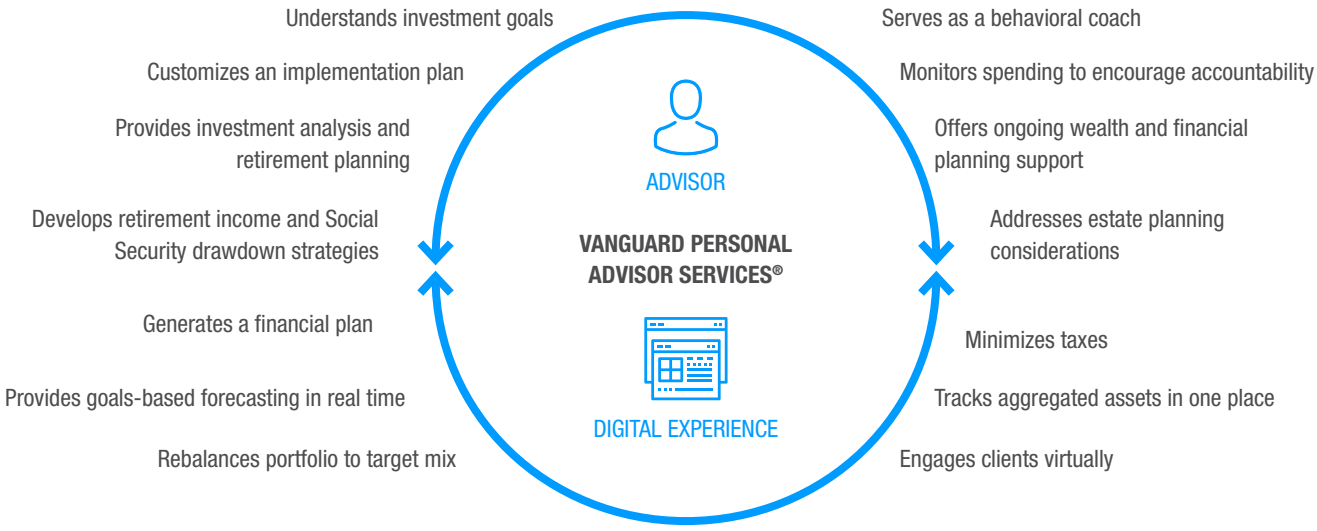
Account aggregation is another key component of automated advice. Many investors have accounts at multiple different financial institutions. Account aggregation allows all accounts to be viewed in one place, and enables advice based on investments across all accounts. VerticalOne and Yodlee (now merged and part of Envestnet) pioneered this approach in 1999 [Fujii et al. (2002)], and now many investment and wealth management firms offer Yodlee's account aggregation capabilities or their internal capabilities.

Automated advice also relies in part on large-scale econometric market models that estimate likely future returns from different asset classes. Most automated advice systems have such a model at their core. Vanguard, for example, says that many of its automated recommendations are based on its Vanguard Capital Markets Model [Kolimago (2017)]. Such models take into account factors like macroeconomic conditions, tax rates, and past returns by asset class.

Modern visual analytics play an important role in automated advice. Since many users of the systems are relatively unsophisticated investors, simple graphic displays are often ideal for that audience. Bar charts, pie charts, line charts, and the like abound. Most automated advice systems issue graphic-intensive quarterly and annual reports.

Finally, while traditional analytical models are widely used in automated advice, some financial firms are beginning to use artificial intelligence as well, and machine learning in particular. Morgan Stanley notes that its "Next Best Action" investment recommendation system is based on machine learning models that match investment opportunities to clients [Davenport and Bean (2017)]. Adam Nash, the CEO of Wealthfront, commented in a blog post [Nash (2016)] that the company's recommendations would increasingly include artificial intelligence capabilities, particularly with regard to actual spending, saving, and investing behaviors by customers: "We're firm believers that artificial intelligence applied to your actual behavior will provide far more powerful advice than what traditional advisors offer today. The reason is quite simple: actions speak louder than words.

Figure 1: Advisor versus machine tasks in Vanguard's Personal Advisor Services



Observed behavior can't be fudged on the phone or lied about in person. More importantly, observed behavior may reveal insights about ourselves that we aren't even consciously aware of."

It seems likely that other financial firms will begin to use machine learning, natural language processing, and other AI capabilities if they aren't already.

4. HOW DOES AUTOMATED INVESTING ADVICE CURRENTLY WORK?

Automated financial advice today is a hybrid process of machine and human participation. The relevant humans may be either the client, an advisor, or both. Vanguard, for example, has a clear division of labor among advisors and machines in its hybrid offering called Personal Advisor Services (Figure 1), which is typical of other hybrid systems.

An important first step (after a contractual agreement has been signed) is for the client to supply information. In most cases this is done directly into the computer. The client provides information on financial goals, family demographics, asset allocation preferences, financial needs, and risk tolerance. Goals most frequently include retirement planning, but may also involve saving for a home, college, or even a car.

After the client data has been supplied, a computer program constructs a proposed portfolio of ETFs and mutual funds, or (less commonly) recommends particular stocks or bonds. In hybrid offerings, there may be a meeting with the advisor to clarify goals and objectives or answer questions. In most cases, the client has several days to agree to the proposed investments. After the client has agreed, the money is invested.

Over time, the machine performs an ongoing and repeated set of tasks, including rebalancing assets, identifying losses for tax loss harvesting, regular reporting, and analytics (including Monte Carlo simulation to show the likelihood of having sufficient funds through a lifetime). The results of account changes are typically displayed to clients on firms' websites.

Hybrid human/machine programs typically feature occasional meetings with advisors. Some, like Morgan Stanley's Next Best Action approach, mediate all recommendations through the advisor. At Vanguard, the Personal Advisor Services offering features advisors as "investing coaches," able to answer investor questions, encourage healthy financial behaviors, and be, in Vanguard's words, "emotional circuit breakers" to keep investors on their plans [Bennyhoff and Kinnery (2016)]. The PAS approach has been highly successful, quickly gathering more than U.S.$80 billion in assets under management – far more than any other U.S. firm thus far.

Some firms that initially offered machine-only services have now moved to incorporate some human contact. Betterment, for example, offers two plans (with higher

fees) that include either annual or unlimited phone consultations with advisors. Personal Capital is also a hybrid service. Wealthfront, however, maintains its machine-only approach to advising.

Whether hybrid or machine-only, all automated advisors offer lower advising fees than purely human advisors. Automated advice generally costs between 0.2% and 0.5% of the client's assets, versus 1.0% or more for human-advised investing [ValuePenguin (2017)]. Some firms have tiered rates depending upon how much human advisor contact is allowed, or the amount of client assets under management.

Automated wealth or asset management also typically requires lower minimum balances for investors than human-only offerings. At Vanguard, for example, the minimum investment level for its human-advised asset management services was U.S.$500,000. But with Personal Advisor Services, its hybrid machine/human offering, the minimum balance is U.S.$50,000. Some online-only services have minimums of U.S.$500 (Wealthfront) or even U.S.$0 (Betterment) [Rieman (2017)].

5. WHAT IS THE FUTURE OF AUTOMATED INVESTING ADVICE?

Other than rapid growth, there are several likely attributes of "robo-advisors" of the future, including the following four domains for change.

5.1 Greater breadth of advice

This is perhaps the best bet for future development. Current versions of automated financial advice are relatively narrow in scope. They address only a relatively small part of consumers' financial lives – investing – and typically only recommend certain types of investments (mutual funds or ETFs).

More advanced investing features would enable investing in different asset classes like real estate, precious metals, or oil and gas. The systems could also focus on tax efficiency and optimization, the management of trusts, IRA management, 401K management for businesses, and so forth. One investment company estimated that there were 115 possible asset classes, but their existing robo-advisor only dealt with ten percent of them.

Automated advice will also extend into areas of financial services beyond investments. Robo-advisors are already also used in insurance to provide automated advice [Schneeweiss (2017)]. Startups like Lemonade and Insurify are using artificial intelligence to engage in chat with customers and evaluate claims. They also have algorithms to recommend levels and types of coverage. USAA, an insurance and banking firm for U.S. military veterans, has a robo-advisor that provides advice not only on insurance, but also investing and spending [Gipson (2015)].

There will also be automated solutions aimed at the financial needs of particular customer segments. Wealthsimple, for example, a Canadian robo-advisor firm, offers systems for both socially responsible investing and Shariah-compliant investing for Muslim customers [George-Cosh (2017)].

5.2 Increased focus on risk mitigation

Most automated systems are not very transparent in terms of how they invest customers' money. The algorithms that they use to select investments or identify customers' risk tolerances, for example, are rarely publicized or made available. Although investment advisors have fiduciary responsibilities, it is often difficult even for regulators to prove that the systems' recommendations are in the best interests of customers. In addition, there may be operational (e.g., trade execution), security, and technical risks associated with automated advice systems.

While few customers appear to be concerned by these risks, regulators (the SEC and FINRA in the U.S., for example) have already issued rulings that specify that the risks are being addressed. And some accounting firms are beginning to offer services to assess algorithms, rules, and other system components to ensure that they do what they say they do, and that unnecessary risks are not incurred [Ameel and Stephenson (2017)].

5.3 New investing models

Almost all automated investment advice is based on so-called "modern portfolio theory," first published by Harry Markowitz (for which he won a Nobel Prize in Economics) in 1952 [Markowitz (1952)]. This theory requires the advisor or system to ascertain the investor's risk tolerance, and then a set of asset classes (theoretically uncorrelated) are assembled in relatively fixed percentages to create an "efficient frontier" portfolio with optimal expected investment returns.

But modern portfolio theory is not the only way to construct a portfolio. Today, there are multiple

alternatives to it, including approaches based on behavioral finance, those that incorporate alternative asset classes, and those that allow for tactical asset allocation, or more flexible allocations over time.

As robo-advising technologies become more intelligent, they will increasingly be able to adopt some of these emerging strategies. Almost all of them would require more data and more calculations than the existing generation of automated advisors. Much of the external financial data is already available and is being employed by sophisticated professional investors.

5.4 Better customer knowledge from data

The more knowledge a financial advisor has of customers' financial behaviors, the better the recommendations can be about how to improve their financial situations. But to expect customers to supply extensive data is a burdensome customer experience.

In other financial sectors, firms are increasingly using external data to learn more about customers and minimize their data entry burden. Home insurers, for example, are employing satellite imagery of homes, rather than having to climb on the roof to inspect it. Automobile insurers are allowing customers to take smartphone photos of accident damage, rather than traveling to a claims center. Some providers of commercial loans, including Kabbage, are given permission by customers to connect to their Ebay or Paypal accounts, Amazon.com sales data, Intuit Quickbooks data, and so forth.

In the future, it is likely that automated financial advisors will also be able to connect to multiple sources of data in order to provide better recommendations. Access to a bank account or to credit card statements, for example, would give a robo-advisor an excellent window into a customer's earning and spending habits. The forays by several advisor firms into account aggregation, and the move by Wealthfront into using artificial intelligence to monitor customer investing behaviors, are just the beginning of this trend.

Of course, this external data access will have to be done with the permission of the customer to minimize privacy concerns. And advisor firms will have to be careful not to use the data for any purposes other than those that truly benefit the customer.

5.5 More market knowledge

Robo-advisors have thus far been largely based on passive investing and "set and forget" portfolios. But they don't have to be. More sophisticated technologies could take into account moment-by-moment market moves and changes in desirability of particular investments or asset classes. Again, this strategy has already been adopted by investment banking trading desks and hedge funds, and it seems likely to "trickle down" over time to individual investors' portfolios. And the vast amount of data and short timeframes involved require that decisions and actions be made by intelligent machines, rather than by human advisors or retail investors. As Kishnan (2017) put it: "Eventually algorithms and artificial intelligence will take over most aspects of money management, particularly picking investments for clients and for trading."

This approach is already being used in some form by several wealth management firms for their internal use or advisor-mediated work with clients. RAGE Frameworks, for example, a company recently acquired by Genpact, is introducing an "active advising" module in its wealth management software that is used by several leading firms [RAGE Frameworks (2016)]. It includes configurable "intelligent agents" to assist advisors in executing their strategies, advise them of patterns in market and customer data, and continuously monitor for changes in the external environment or the client's personal situation that can impact client portfolios. These capabilities are not yet available to retail investors, but probably will be in the near-term future.

All of these future directions tend to involve more sophisticated and complex investment strategies and technologies. However, key factors in the success of robo-advisors with financial consumers is that they are relatively easy to use and understand, and that fees are kept low. Firms that add these sophisticated features will have to balance their complexity with these other objectives.

6. WHAT ARE THE IMPLICATIONS FOR HUMAN INVESTMENT ADVISORS?

While many of the investment-picking aspects of financial advising will undoubtedly be taken over by machines, there are still some important roles for humans. Perhaps some advisors will lose their jobs, but probably not in large numbers. What roles can advisors continue to play?

There are a variety of possibilities for roles working alongside smart machines in various fields [Davenport and Kirby (2016)]. People who were formerly traditional financial advisors could become experts, for example, in how robo-advisors work, their strengths and weaknesses, and which advisors are best for particular circumstances. They could be integrators of different online advice sources, and help clients and investment firms to understand what systems to use for what purposes. They could also, like hedge fund managers, analyze the results from machine-advised financial portfolios and assess whether changes are necessary in the algorithms and logic employed by the machines.

Advisors could also shift to providing advice on investing for relatively obscure asset classes that are not included in automated advice systems. An advisor who specialized, for example, in distressed debt investing or assets like timber or oil and gas exploration would be unlikely to be replaced by a machine anytime soon.

Perhaps the most common role for financial advisors in adding value to smart machines is behavioral coaching. Over the past decade, many academics and investment firms have come to realize that behavioral and psychological issues play a large role in investing [Montier (2007)]. Deciding what investments to buy, and when to buy and sell certain investments, are often not entirely rational processes. As investment selection is taken over by algorithms and artificial intelligence, coaching investors on the appropriate behaviors for their situations can be a valuable role. Behavioral coaches could, for example, dissuade clients from buying at the top of the market or selling when markets crash. They could attempt to reconcile the diverse risk perspectives of husbands and wives who are investing jointly.

Several investment firms that have made substantial commitments to automated advice have embraced behavioral finance and coaching. Primarily online firms like Betterment and Wealthfront have included materials about behavioral finance on their websites. And firms with hybrid machine/human advice offerings, like Vanguard's Personal Advisor Services, have encouraged financial advisors to learn more about behavioral coaching and to play that role with clients. Vanguard also makes extensive use of video interactions between advisors and clients to try to increase the engagement level of coaching interactions [Kolimago (2017)].

Of course, some investors will continue to prefer human advice, particularly at the high end of wealth management and for older clients. Hence, some advisors will not be greatly affected by automated advice, at least over the next decade or so.

However, many advisors will feel an impact from the robo phenomenon. As in other fields, financial advisors who want to keep their jobs may have to be flexible and adaptive. They may have to learn new skills in terms of understanding financial technologies, or in terms of mastering behavioral coaching. They may have to change their asset focus, or modify their business model. However, those who are willing to make such changes are likely to remain employed.

For the firms that employ those advisors, automated advice is likely to have mixed implications. Fees for advising clients and managing portfolios are likely to drop, as they already have at firms that have aggressively adopted robo-advice or hybrid machine/human models. However, the combination of high-quality automated advice at a relatively low cost could bring large volumes of new clients into the market for investing advice. Firms catering to the "mass affluent" [Nunes and Johnson (2004)] market are most likely to benefit from this market growth.

While the details of adoption of automated financial advice are unclear, there is little doubt that it will become increasingly popular. Financial services firms, financial advisors, and clients will all see substantial change in the financial advice process over the next several years. Extended face-to-face discussions between client and advisor may not vanish altogether, but they may become endangered.

References

Ameel, S., and C. Stephenson, 2017, "Hands off-guard up: for wealth and asset management firms, "robo-advising" platforms carry new risks," Deloitte, http://bit.ly/2xX7rNM

Bennyhoff, D. G., and F. M. Kinnery, 2016, "Vanguard Advisors' Alpha," Vanguard, June, http://bit.ly/2gXMDCs

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

Davenport, T. H., and R. Bean, 2017, "How machine learning is helping Morgan Stanley better understand client needs," Harvard Business Review, August 3, http://bit.ly/2u8z8pk

Ellis, C. D., 2017, "The end of active investing?" Financial Times, January 20, http://on.ft.com/2y1FwwX

Epperson, T., B. Hedges, U. Singh, M. Gabel, 2015, "Hype vs. reality: the coming waves of robo-adoption," A.T. Kearney Report, June, http://bit.ly/2vRLuiu

Europe Economics, 2014, "Retail distribution review post-implementation review," December 16, http://bit.ly/2y1Gr0n

Fujii, H., T. Okano, S. Madnick, and M. Siegel, 2002, "E-aggregation: the present and future of online financial services in Asia-Pacific," MIT Composite Information Systems Laboratory Working Paper, September, http://bit.ly/2h0sXKV

Fuller, J., and J. Patrie, 2017, "Fitch: robo-advisory could mark shift for U.S. wealth managers," FitchRatings press release, July 12, http://bit.ly/2wXfJWM

George-Cosh, D., 2017, "Robo-adviser tries to reach Muslim investors," The Wall Street Journal, August 24, http://on.wsj.com/2wXhyo1

Gipson, Z., 2015, personal interview with the author, USAA's Chief Innovation Officer, August 5

Kishnan, S., 2017, "Robots are eating money managers' lunch," Bloomberg BusinessWeek, June 26, https://bloom.bg/2rPYXom

Kolimago, F., 2017, personal interview with the author, head of Vanguard's Personal Advisor Services, July 12

Mackintosh, J., 2017, "Robotic hogwash: artificial intelligence will not take over Wall Street," The Wall Street Journal, July 17, http://on.wsj.com/2h0cE4d

Markowitz, H., 1952, "Portfolio selection," Journal of Finance 7:1, 77-91

Montier, J., 2007, Behavioural investing: a practitioners' guide to applying behavioural finance, Wiley

Nash, A., 2016, "Introducing Wealthfront 3.0," blog post, March 31, https://wlth.fr/2f2ljBr

Novick, B., B. Lu, T. Fortin, S. Hafizi, M. Parkes, and R. Barry, 2016, "Digital investment advice: robo advisors come of age," September, http://bit.ly/2mgglvY

Nunes, P., and B. Johnson, 2004, Mass affluence, Harvard Business Review Press

RAGE Frameworks, 2016, "RAGE launches LiveWealth Solution featuring cognitive intelligence for active advising," press release, January 11, http://bit.ly/2y1xm83

Reuters, 2017, "Fitch: robo-advisory could mark shift for US wealth managers," July 12, http://reut.rs/2xhJfZl

Rieman, M., 2017, "Which online advisor is best for you?" Nerdwallet, https://nerd.me/2wWVdHe

Schneeweiss, Z., 2017, "Robo-advisors gaining as insurance sales agents," Insurance Journal, http://bit.ly/2xhCqY1

Srinivas, V., and U. Goradia, 2015, "The future of wealth in the United States," Deloitte Center for Financial Services, November 9, http://bit.ly/2hVklEj

Tergensen, A., 2017, "Fiduciary care from a robo-adviser?" The Wall Street Journal, February 10, http://on.wsj.com/2wXvXRv

ThinkAdvisor, 2015, "William Sharpe—the first robo," ThinkAdvisor, http://bit.ly/2wXiEA9

ValuePenguin, 2017, "Robo-advisor fee comparison," ValuePenguin, http://bit.ly/2xtL32D

Weil, D., 2017, "Inside the decades-long quest to bring down financial information giant Bloomberg," Business Insider, July 18, http://read.bi/2y1JtBU

Understanding robotic process automation (RPA)

MARKUS ALBERTH | Managing Principal, Capco
MICHAEL MATTERN | Managing Principal, Capco

ABSTRACT

Robotic process automation (RPA) is the use of software as "virtualized workforce" to operate applications like a person processing a transaction or completing a process in front of a computer screen. Currently, the logic is still mainly rule-based and robots can relieve workers to do routine process work. In the near future, artificial intelligence will enable software robots to automate more and more work of humans with respective social and financial implications. With already over 50 providers, the market gains speed and volume, and innovation will lead to increasingly fragmented segments. In light of related and competing automation approaches, RPA has prominent benefits and typical caveats, such as quick and predictable cost cuts and scalable near real-time service potential. They translate into elements of a potential business case. Real business scenarios suitable for RPA show that it is working. The implementation of an RPA solution has at least three phases, the proof of concept, the pilot, and the leverage phase to other use cases within the company. If not done well, robots may be too slow, too expensive, and introduce too much complexity. A neutral partner with professional knowhow can neutralize these risks. In addition, the benefits of RPA can most probably be harvested earlier; making it right from the beginning. Whatever you think about robots: robots and RPA are here to stay. Robots are cheap and best in data processing, consequently, they will impact the respective processes along the value chain of a lot of industries for the benefit of the company and the customer.

1. INTRODUCTION

Robotic process automation (RPA) is the use of software as "virtualized FTE" to operate applications like a person processing a transaction or completing a process in front of a computer screen. To accomplish this, robots use their own functional user IDs to log in and out of the operated applications. "Macros" are a kind of early stage in the development history of software robots.

RPA, therefore, does not replace existing applications or manipulate their code, but rather works with those systems in a manner similar to a human user. Some robots replace approximately one worker, some replace up to five workers. Nevertheless, RPA is currently not yet able to fully replace human work. Only simple, predictable tasks can be automated, while more sophisticated work is still left for human subject matter experts.

RPA is currently guided by rules rather than artificial intelligence (AI). It allows for escalation to a human supervisor in cases where the ruleset does not contain a suitable response for a specific situation. However, in the future, AI will be increasingly integrated within software robots to take over more human tasks.

RPA solutions create an audit trail for every action taken by the "virtual FTE" so that compliance to process guidelines can always be proven after the fact.

2. MARKET FOR RPA

The market for RPA solutions has developed rapidly. According to Forrester [Le Clair et al (2017)], there are more than 50 providers in the market with prices ranging between U.S.$ 5,000 and U.S.$ 10,000 per robot, depending on provider and functionality. According to a 28-criteria evaluation, Forrester classifies the following providers as leaders: Automation Anywhere, Blue Prism, and UiPath; as strong performers with competitive options: WorkFusion, Pegasystems, NICE, Kryon, EdgeVerve, and Redwood; and as contenders: Kofax, Contextor, and Softomotive.

The current focus of innovation is on robot management and governance functionality, such as central control of robots, preservation of formerly human process knowledge, and governance, such as connectivity monitoring, rollback capabilities for processing failures, and testing capabilities for application changes. AI seems to be the next focus; probably in a few years. This means that robots are no longer mere commodities. They are growing in functional maturity, which means that potential users should look at the offers in detail rather than rush to sign with the first provider they meet. The same is true for IT service companies that offer implementation and other services surrounding RPA.

3. DEVELOPMENT OF RPA SOLUTIONS

The early incarnations of today's RPA were mainly screen scraping solutions that sought to integrate new software applications with legacy applications that had no readily available means for automated interfacing.

Evolving to rule-based machines that could automate business processes across system and organizational boundaries, RPA solutions have recently started to leverage machine-learning approaches to improve process quality over time, alleviate recurring situations that require human intervention, and generate new insights from legacy application landscapes.

> "Right now, many finance jobs require people to act like robots, so they'll easily be replaced by robots."
>
> [Rosenfeld (2017)]

While previous solutions required human interaction whenever none of the predefined rules applied to the case at hand, machine learning enables future RPA solution to learn from how exceptions are handled by human operators to eventually enhance their ruleset. Machine learning is also applied to classify unstructured documents, such as Kofax Kapow and WorkFusion Intelligent Automation, in preparation for RPA processing or assist in building or improving the rule-base for an RPA deployment, such as Automation Anywhere IQ Bots.

Intelligent virtual assistants, such as IPSoft's Amelia, aim to bridge the gap between customers communicating via natural language and highly automated back-end processes. WorkFusion blends machine-based robots with crowdsourced humans to enable automation of process steps that exist at the boundary between "machine work" and "expert work"; still out of reach for a pure robotic solution but standardized enough to be distributed among machine and low-skilled human workers. Overall, the impact of further developments in machine learning, automation, and AI put a large

percentage of human employment at risk. Frey & Osborne (2013) estimate, for example, that about 47% of total U.S. employees have to be considered at a high risk of being replaced by machines in the foreseeable future (about one to two decades). As Deutsche Bank CEO, John Cryan, stated recently: “Right now, many finance jobs require people to act like robots, so they’ll easily be replaced by robots” [Rosenfeld (2017)].

This will have an impact on the financing of social security systems, which are usually connected to the direct income of the workforce. Microsoft founder, Bill Gates, and Siemens CEO, Joe Kaeser, have already proposed to tax software and robots to solve that issue.

4. RELATED AND COMPETING APPROACHES

Unlike competing technical strategic approaches, such as EAI (enterprise application integration), SOA (service-oriented architecture) with enterprise integration layer, or business process management with automation functions, the more tactical RPA does not require changes to existing applications and thus does not trigger any larger IT change projects. Typical payback periods for investments in an RPA initiative are measured in months rather than years.

Unlike competing cost cutting approaches, such as outsourcing/offshoring/nearshoring or any other labor cost arbitrage like in-house outsourcing to legal entities with a cheaper labor tariff structure, RPA keeps everything in-house and onshore. In addition, RPA also benefits from other factors, such as higher process quality.

Other well-known benefits of RPA are:

- Quick productivity gains within weeks, or a few months, i.e., almost instant cost cuts.
- Upfront investment costs and license fees are small and can be calculated reliably, as can the return on investment (RoI).
- Suitable as a tactical interim cost cutting solution if strategic technical solution is still too expensive.
- Robots can work 24 hours a day, 7 days a week.
- No, or minimal, process changes are needed, though the introduction of RPA can trigger process improvements.
- No, or minimal, application changes required.
- It is scalable and benefits from economies of scale.
- Improved quality output compared to human workers, i.e., lower failure rate and risks.
- Continuous and transparent compliance documentation.

- Potential for process improvements during implementation, because of detection of shortfalls, gaps, etc.
- Lean Six Sigma programs can benefit from introducing RPA for highly standardized tasks, since process repeatability of a "virtual FTE" produces a lot of data, which is required for six sigma, and removes humans as possible sources for errors.

The advantages are clear and simple; the potential caveats are not. Nevertheless, they need to be considered for sound management decisions on the use of RPA and on the selection of the right provider.

Typical caveats connected to RPA initiatives comprise:

- New IT architecture feature (the robots) need to be serviced rather than a strategic layer integration.
- New IT systems (the robots) require new IT security coverage.
- Reduces business case for strategic solution and, therefore, may delay the strategic solution.
- Currently still for routine work only, i.e., standard processes need to be cut out of the end-to-end process logic to be automated by RPA.
- The resulting fragmented part of the end-to-end process still needs to be serviced by human workers. Depending on the individual process management layout, this can increase process complexity for human workers. This can mean more setup times at the interfaces between RPA-process parts and human worker process parts, which could mean more failures on the human side.
- Currently for paperless work only; data needs to be digitized.
- Once automated, processes are out of sight and can, therefore, shift out of focus for process improvements.
- RPA is just this – automation. It does not trigger or replace the surge for new digital business models. RPA should not distract you from that task.
- Robots need to be supervised; the work does not just vanish. New tasks emerge with RPA.
- Legal issues may emerge if functional user IDs of robots are misused.
- Social impact of RPA implementation on workforce needs to be taken into account.

5. BUSINESS CASE ELEMENTS

One of the main arguments for RPA is instant cost reduction with small upfront investments and reliable RoI estimations. Here are some main elements of any RPA business case to show the total costs and benefits of RPA ownership.

The benefit side comprises:

- Reduced processing workforce (in euros per year): the number of reduced full-time equivalent capacities depends on (1) the automated process steps and the respective handling time (the longer the handling time, the better; and you have to offset new complexities and fixing costs for additional failures, if applicable), (2) the quality assurance process steps (and thus handling time) formerly needed to ensure the quality of the then automated steps (this also means reduced worktime costs for failure fixes), and (3) the number of transactions per process (the higher the transaction volume, the better).
- Reduced office space (in euros per year): this should have a very tight correlation to the reduced workforce and includes the rent for the office, depreciation of furniture, front-end IT installation and operation (helpdesk, etc.), canteen, etc.
- Reduced costs due to defective processing (in euros per year): clients hold you liable due to defective processing. These costs will be reduced with higher quality without human failure.
- Reduced "FTE overhang costs" or improved scalability/turnover (in euros per year): in the old model, you had to hire new people with increasing turnover volume and lay them off with decreasing turnover volume. (a) If you have a craftsman/artisan shop that is operated by deeply skilled and trained workers, you have to pay for the skill on the workforce market or you have to train them for some 1-3 years. This model does not allow you to "breathe with your costs with the market." Instead you have a time lag in adjusting workforce demand and supply (i.e., your capacity). Because market demand changes, you systematically have either too many (salary costs) or too few (that costs you in terms of turnover and thus income and/or market share) people on board. RPA can overcome those costs because robots are scalable. (b) If you don't have a craftsman/artisan shop but already have an industrialized factory model in place with lower skilled workers that can be 100%

productive within weeks, RPA does not provide you with that benefit, because you can already "breathe with the market."

- New revenue sources because of new products (near real-time; in euros per year): robots work 7x24 and are just-in-time scalable, depending on their buffer capacity. This makes it possible to have very short service level agreements, which in turn allows for new offers to clients or new client experiences. This can lower the cost of client acquisition and improve client retention and opens the doors for new income sources.

The cost side comprises:

- Limited one-off investment costs upfront for framework (in euros): costs for internal resources (e.g., to adjust policies, to make decisions) and costs for consultants (e.g., for feasibility study, etc.).
- Limited one-off investment costs upfront per use case (in euros): costs for internal resources (e.g., for technical implementation in the data center), costs for robot provider, costs for service provider, costs for consultants (e.g., for process preparation), and human resource costs (e.g., early retirement costs).
- License costs (in euros per year): typically robots are not bought but rented or licensed. You pay per time unit or per transaction. Very often, there is a minimum time period in the contract, e.g., three years, to cover the total expenses of the provider. These license costs cover upgrades to new versions, helpdesk, hot fixes, etc.
- New workforce costs to control and govern RPA (in euros per year): the cost structure depends on your framework settings. You can decide, for example, on a central control team or several decentral control teams. Front-end changes of applications need to be governed and "trained" for the robots, depending on your product.
- Costs of new workforce to control and govern RPA (in euros): the profiles needed are quite different form the automated process operation profiles and are usually more expensive. You can acquire those skills through training (training costs) or by hiring people (recruiting costs).
- Office space costs for those people (in euros per year): see office space costs on benefit side.

6. BUSINESS SCENARIOS SUITABLE FOR RPA

RPA should be considered in the following situations: need to improve/automate currently manual processes, need to increase FTE productivity while maintaining accuracy, failure to realize RoI on EAI, improve return on BPO initiatives by automating the lower-level tasks instead of just moving to lower-cost human FTE, and new online or mobile front-end desired for legacy back-end architecture.

Tasks suited for RPA include: data entry and validation, file and data manipulation, formatting, and multi-system data entry/reconciliation.

Examples of successful applications of RPA include Bloomberg's use of WorkFusion to automate data capture and crowdsource quality assurance for a base of about 500,000 existing company records that have to be maintained from SEC filing data, or the Co-operative Bank using BluePrism to automate part of their payment processing, deciding whether to process or return payments from accounts with low or insufficient funds.

Hitting a cost-saving roadblock after using traditional labor-arbitrage for their IT processes, a major financial services firm leveraged IPsoft automation to replace labor and vastly reduce manual interventions, resulting in a 35% FTE reduction in support teams during the first year. Finally, USAA used IBM Watson to automate customer services for 150,000 users in a one-stop shop for veterans in need of assistance on matters from job searches to government benefits.

7. IMPLEMENTATION PROJECT

The implementation of an RPA solution has at least three phases, the proof of concept, the pilot, and the leverage phase to other use cases within the company.

The **proof of concept phase** starts with the determination of the purpose of an RPA implementation and identification of potential use cases inside the company. Set out your objectives for an RPA program and take those objectives as initial objectives only, since you can be in danger of simply following the herd, without fully grasping the consequences. Confirm and adjust those objectives during the project. It is essential to have those objectives to measure their delivery later on.

The next step is to look for use cases. To find them, you have to look at end-to-end processes, as well as their details. Depending on your process landscape, you might find that only parts of end-the-end processes may be suitable for RPA automation. Scan your total process landscape systematically to find all and the most profitable use cases. A typical mistake is to identify the first use cases "by coincidence." To exploit your entire benefit potential of RPA usage you need a systematic approach.

Look out for RPA providers and/or RPA service providers that best support your objectives and use cases. Get familiar with the products. Select your shortlist of providers, and use the data to draft your first business cases.

It is at this stage that you are able to build on your internal RPA framework. Most probably, you have to adjust policies and/or write new ones. Further, you have to make technical decisions like introduction of a functional user ID, architectural "location" of the robots, etc.

Some important questions that need to be answered are: who will be responsible for RPA governance: business or IT? Who will control them during daily production: backoffice, data center production, or a new group wide central RPA control group? What adjustments are needed to change the management of IT access rights? Who will apply for the access rights of a new robot? Does the granularity of the access rights fit for the new robot use cases? What business lines will be involved? What access will audit and compliance get? And, what are the new performance measures?

At the end of the proof of concept phase you will have objectives, use cases, a selected provider, or at least a shortlist, and your internal framework preparation completed. In addition, a first rollout roadmap could be drawn.

The **pilot phase** focuses on the implementation of one or few RPA use cases. You start with the preparation of the procedural and technical environment.

Paper needs to be transformed into electronic data (e.g., via scan, OCR, (free-)form capturing) or flawless dataflows without data breaches need to be introduced, using, for example, client front ends that provide the data electronically instead in the form of paper. However, that is not the end of the story. Even the electronic data need to be standardized for RPA use; RPA may require the data to be cleaned. If, for example, some records provide data content in one data field and others don't, the robot may be confused and stop working or it will produce nonsense process results.

An electronic process trigger needs to be implemented to initiate an RPA transaction later on. The robots need to be technically rolled out in the data center (or somewhere else) and customized for your special need. They also need to be equipped with the functional User ID.

The end-to-end processes have to be cut according to the future allocation of work between robots and humans, and the former needs to learn the routine processes that it will perform in the future.

Finally, the robots need to be tested.

With these preparations, the pilot itself can be initiated. It needs to be monitored to deliver the information needed to optimize the RPA operation according to your objectives and to adjust the rollout roadmap.

You have now identified the essentials necessary for leveraging the RPA concept across all of your use cases and to even identify more of them in your company. This leveraging phase will hopefully deliver on your objectives to implement RPA.

To implement **RPA as a sustainable tool of your management system**, you should expand your framework by a process to systematically detect all use cases in your environment that can potentially be automated by RPA, if you have not done that already in the proof of concept phase. In fact, you have to assess all manual work in the backoffice – and yes, also in the front office. Very often, there are still a lot of administrative tasks in the front office that can be done by RPA. And here we are not talking about client facing robotics. Unproductive, but necessary administration, might be everywhere.

Typical project risks: If you don't plan the RPA introduction well, robots may be too slow, too expensive, and introduce too much complexity. Some cases show that the implemented robots could not be used at all. A neutral partner with professional knowhow can neutralize these risks. In addition, the benefits of RPA can most probably be harvested earlier; making it right from the beginning.

8. OUTLOOK

Robots and RPA are here to stay and promising to gradually become smarter over time as vendors compete to increasingly include machine learning into their solutions.

You can use this new technology to automate your old business model or you can utilize it for a new digital business model.

Robots and human workers will work side by side, with each focusing on their competitive edge. Robots are best in data processing, consequently, they will impact the respective processes along the value chain of a lot of industries for the benefit of the company and the customer. The **right use** of technology is the key to sustainable success. This could be the value add of consultants, who understand you, your business, the future of your business, and the new technology of software robotics.

References

Frey, C. B., and M. A. Osborne, 2013, "The future of employment: how susceptible are jobs to computerization?" report, Oxford Martin School, University of Oxford, http://bit.ly/2yym1vt

Greer, S., and C. Beattie, 2016, "Robo op - how robotic process automation is applied in banking operations," Celent, November 28, http://bit.ly/2wNThzb

Le Clair, C., A. Cullen, and M. King, 2017, "The Forrester WaveTM: robotic process automation, Q1 2017," Forrester, February 13, http://bit.ly/2kaojyt

Rosenfeld, E., 2017, "Deutsche Bank CEO gets brutally honest about what automation is going to do to banking jobs," CNBC.com, September 17, http://bit.ly/2xH0zXu

Rürup, B., 2017, "Abgabenbescheide für roboter," Handelsblatt, May 2, p.12, http://bit.ly/2fXHDL5

Tornbohm, C., 2016, "Market guide for robotic process automation," Gartner, November 7, http://gtnr.it/2nV6XG8

Robotizing Global Financial Shared Services at Royal DSM

MARY LACITY | Curators' Distinguished Professor, University of Missouri-St. Louis, and Visiting Scholar, MIT CISR
LESLIE WILLCOCKS | Professor of Technology Work and Globalization, Department of Management, The London School of Economics and Political Science
ANDREW CRAIG | Associate Researcher, The Outsourcing Unit, The London School of Economics and Political Science

ABSTRACT

Robotic process automation (RPA) is a new breed of software robots designed to automate services through processing structured data by following rules. The research examines how RPA is being deployed at Royal DSM, the Dutch multinational company, in its global financial shared services organization. Here, like many other companies, DSM adopted RPA as the next transformation lever beyond centralization, standardization, and relocation to a lower cost region. DSM's first phase of automation "robotized" its financial close processes for three of its global business groups. Phase I happened quickly – within a few months. Phase II scaled the RPA adoption by migrating another six business groups on to the RPA platform. This article details the multi-faceted business results DSM achieved, the crucial role of a suitable software provider, and eight key lessons from the DSM experience that other companies can learn from.

[1] We thank the customers, providers, and advisors who were interviewed and gave so generously of their time for the research program of which this case forms a small part. In particular, we acknowledge and thank all at Redwood Software (Redwood) and Royal DSM who supported us so tirelessly through the multiple stage of the research process for this article.

Recent books by the authors on these themes are *Service automation robots and the future of work*, and *Robotic process automation and risk mitigation: the definitive guide*, available from www.sbpublishing.org.

1. INTRODUCTION

Robotic process automation (RPA) is a form of software that can be applied to structured data, and stable, repetitive, rules-based processes (or sub-processes) that produce deterministic outcomes. However, there are over 39 RPA software providers, and the software tools tend to have different characteristics and uses. The advantages of RPA, and the reasons for dramatic uptake in 2016/17, are the "lightweight IT" aspects – RPA tends to be cheap, quick to implement, and does not lock up IT (information technology) resources. RPA is also designed to be configured by subject matter experts instead of IT programmers, and interfaces with existing systems of record the way a human does, through the application user interface.

including FTE (full-time equivalent) savings, increased service quality (because software robots execute exactly as configured to do), increased service delivery speed, and redeploying human talents to more challenging work. These business benefits, however, can only be achieved with proper governance. DSM epitomizes the emerging recognized best practices for achieving business benefits. These best practices include senior management support, control by business operations/ shared services, talent redevelopment, and change management to prepare the organization for changes caused by automation [Lacity and Willcocks (2015, 2016a, b), Lacity et al. (2015), Willcocks and Lacity (2015a, b, c, d, 2016), Overby (2016)].

Table 1: Royal DSM's RPA capabilities at a glance

	TIME FRAME	PERCENTAGE OF MANUAL TASKS AUTOMATED	NUMBER OF AUTOMATED PROCESSES	SCALE	RESULTS
PHASE 1	July to November 2015	89%	19	Migrated 3 business groups, comprising 60 company codes	• Faster financial close – from over 2 weeks to just 3 days • Increased accuracy and compliance • 45 FTEs freed up for more valuable work • RoI in 9 months
PHASE 2	February to August 2016	89%	25	Migrated 6 business groups, comprising 130 company codes	

The technology we look at in this case is slightly different. To deliver scalability, resilience, and security, Redwood Software's[2] preferred method is to interact with core ERP and other systems through APIs and other standard integration methods. In instances where there are no interfaces, they can use desktop interfaces. To provide end users/subject matter experts, instead of IT programmers, with the easiest method for configuring robots Redwood Software's design allows robots to be fed with business parameters, which are external to the technical robot definitions themselves. These include time zone, period, year, company code, account selection, cost centers, allowed deviations, exchange rate type (month end rate, month average rate), rules for provisions, and any parameter that comprises part of the process that the robot performs.

DSM, like other early adopters we studied, achieved multi-faceted business results from deploying RPA,

In this article, we describe DSM's successful implementation of RPA, using Redwood Software. We compare DSM's practices and challenges with other RPA adoption cases we studied. Companies still considering RPA adoption can learn valuable insights from DSM and from other earlier adopters. To underline the business value achievable with RPA, we begin this article with the end results. In August 2016, DSM was deploying over 60 software "process robots" to automate about 89% of the manual tasks associated with its financial close process (Table 1). DSM earned a positive return on investment (RoI) within nine months. DSM shrank its financial close process from 15 to three days using RPA. Since, then DSM has moved to automating other financial and accounting operations.

[2] Redwood Software is a software company with a mission "to help organizations achieve 'The Robotic Enterprise™'" www.redwood.com

2. ROYAL DSM – THE BUSINESS CONTEXT FOR ROBOTIZATION

To put the RPA journey into context, we explain here DSM's business background. DSM is a multinational company operating in 50 counties and headquartered in Heerlen in the Netherlands. It operates in three industries: health, nutrition, and materials. The company produces vitamins, carotenoids, premixes for food and feeds, enzymes, minerals, cultures and yeasts, pharmaceuticals, bio-plastics, and coating resins, to name but a few of its products. DSM is recognized as an innovator in biomedical materials, advanced biofuels, bio-based chemicals, and solar systems [de Haas (2016)]. Its motto is "bright science, brighter living." In 2015, DSM earned profits of €88 million on €11.5 billion in sales. It employed 20,750 people worldwide, of which 23 percent work in the Netherlands.[4]

In August of 2011, DSM's CEO, Feike Sijbesma, announced the intended changes to DSM's organizational and operating model: "with the aim of creating a more agile, focused and cost-efficient organization, with a stronger business and market focus and globally leveraged support functions."[5]

Part of this strategy included the expansion of shared business services. DSM had already created global shared services for IT and human resources (HR). It next aimed to add financial services.

3. FINANCIAL SHARED SERVICES (FSS) LAUNCHES IN 2012

DSM's FSS journey began in 2012. After considering several sourcing and location options, DSM rejected outsourcing and decided instead to have five regional business service offices and one large captive center in Hyderabad, India. The captive center would create the dual benefits of lower costs through labor arbitrage while still keeping the employees engaged with and connected to DSM. The FSS plan called for all of DSM's business units to be migrated to the financial shared services within four to five years. Why so long? Like most multinational companies, the main challenge was standardization. DSM's history includes many mergers and acquisitions (M&As) that bring along their legacy systems and processes. DSM, for example, had several Enterprise Resource Planning (ERP) platforms, including 13 different SAP implementations.

> "Our purpose is to create brighter lives for people today and generations to come. We use our unique competences in health, nutrition and materials to create solutions that nourish, protect and improve performance."
>
> [DSM 2015 Annual Report][3]

FSS decided to build a seamless gateway on top of the legacy systems rather than bear the expense of implementing one global ERP platform. This quickened implementation and FSS was able to migrate 130 company codes to which robotics were applied across Europe, Asia Pacific (including China), and the U.S. by 2015. In 2015, the CEO requested that FSS speed-up the remaining migrations. A new global improvement program called "Arjuna"[6] was started. The Arjuna project aimed to propel FSS to be in the top quartile of the world's best global shared services performers. DSM had room to improve; a recent benchmark indicated that its credit management was performing at the 45th percentile and its accounts payable and account to report services were performing at the 65th percentile. For example, DSM was taking 15 days for financial close when top performing companies were closing in 3.28 days. FSS's mantra became "F for First time right, S for Simplification, and S for Standardization."

Theo de Haas, Senior Business Partner Group Services for FSS, and his senior staff examined the practices used by top performers. They attended seminars and increasingly heard that top performers do three things [de Haas (2016)]:

1. They make greater use of process and technology-related best practices.
2. They move non-essential activities out of the critical path so they can be managed and resolved during the month.
3. They automate many traditional clerical manual tasks with RPA.

In particular, RPA was being touted as a new breed of software robots, designed to be used by subject matter experts, and that it interacted with existing systems of

[3] Source: 2015 DSM Annual Report, http://bit.ly/2uPTYan
[4] https://en.wikipedia.org/wiki/DSM_(company)
[5] DSM 2011 Annual Report available, http://bit.ly/2vkGlgY
[6] "Arjuna" is a Hindu word that translates into "ruler," "one who guides," or "one who is not un-victorious." It is also the name of the third Pandava brother in Hinduism (source: https://en.wikipedia.org/wiki/Arjuna).

record. But the FSS leadership had many questions: is RPA secure? How long do RPA implementations take? How will RPA interact with the ERP systems? De Haas decided to assess the RPA concept, thus beginning DSM's robotization journey.

4. ROYAL DSM'S ROBOTIZATION JOURNEY

Like most organizations, FSS began their RPA journey with a proof of concept (PoC). This aimed to assess the financial and technical feasibility of RPA. The FSS leadership decided to examine the financial close process as the test case. Would automation produce business benefits? The team in FSS's Indian captive center examined the number of manual steps it took for the period end close process. There were an astounding 485,000 manual activities per month in the financial close process. Humans were not only doing the transactions, but also pausing to document each step to ensure compliance. de Haas and his team estimated that they could easily automate 60% of the manual tasks and achieve the target RoI. Furthermore, the quality would improve because the software robots would follow all the rules.

Convinced of its financial value, de Haas approached the Chief Financial Officer (CFO) with the RPA proposal. According to de Haas, the CFO asked him how he would guarantee the close. He recalled her saying: "Don't come back in two months and say to me, 'we have the month-end figures, but don't ask us if the figures are correct because we don't know anymore because the robots are doing it.'"

Hence, any RPA solution would require built-in controls and checkpoints to verify that the process ran correctly. FSS decided to build a test solution and turned to its existing software provider, Redwood Software, for help.

5. FSS SELECTS REDWOOD SOFTWARE AS ITS RPA PROVIDER

Redwood was an obvious fit with DSM. DSM's FSS was already using SAP Financial Closing Cockpit and knew that Redwood already helped to automate many of this system's background tasks. FSS also knew that Redwood offered RPA tools. FSS valued that Redwood was focused on financial services and had deep subject matter expertise. Redwood Software's sales team understood in detail how journal entries, reconciliations, and other financial processes work: *"With Redwood, we talked to financial people. We did not talk to software people. Redwood knows how financial processes work."* – Theo de Haas, Senior Business Partner Group Services, FSS.

In contrast, according to Theo de Haas, many other RPA providers pitched the technical capabilities of their products instead of their ability to optimize financial processes. Redwood is also fully implemented with SAP – a big advantage from FSS's perspective given that they were running global financial services on 13 versions of SAP. Redwood's VP of Worldwide Marketing, Simon Shah, confirmed the DSM analysis: *"What seemed to resonate with DSM is that we spoke both the language of robotics and the language of finance. Our robots communicated directly with ERP, while shielding business users from the technicalities – all within a much larger strategic framework from the outset."*

The technical feasibility of RPA was tested on the month end close process for one of DSM's business groups, Engineering Plastics. The PoC team replicated a previous month's end close using Redwood Software's RoboFinance® solution. The software robots executed the business rules and monitored and documented each step in the process. Engineering Plastics confirmed that the software robots produced the exact same figures, but did so much quicker. According to Mohammad-Sajjad Hussain, Lead Business Process Expert for DSM Business Services India: *"We worked closely with Engineering Plastics. Since it was the first time, we gave them a lot of support so in the end they didn't feel that this change was out of control. They were quite satisfied. They didn't see any difference in the ways of working after automation. They didn't see any side effects or have to do a lot of corrections or extra work because of automation."*

The PoC team proved the financial and technical feasibility of RPA. FSS launched a two-phased implementation plan.

PHASE I: JULY TO NOVEMBER 2015

Phase I aimed to automate financial close for three business groups. DSM created an RPA project team that comprised FSS and Redwood employees. FSS assigned three people: two business process experts and a program manager. Redwood assigned four people: three technical software experts and a financial expert. The financial expert from the Redwood team was to examine the proposed process redesign to be sure it would follow best practices and could be performed by the software robots.

Much of the initial work focused on documenting tasks performed by humans in enough detail to be specified as rules for the software. Humans can execute tasks with less detailed instructions than software robots because humans know how to fill in the gaps in instructions. For example, where there is an intercompany imbalance during a reconciliation process, from experience a user might instinctively know which companies have been incorrectly posted to. A robot, however, will have to be configured to trawl through all the companies to locate the incorrectly posted item. Despite this, the robot will process at a much faster speed than the user. Hussain explained: *"Existing documents are for users who unconsciously perform the activities and often they don't even refer to them. But when you are trying to implement robotics, you are asking a robot to do a task…it will not understand based on documents [designed for humans]. So your documentation should be as detailed as possible, and of course, [embody] a clear understanding of what would be the impact before the task and after the task."*

The RPA team also had to redesign the process for automation so that the robots and humans were not constantly passing steps to one another. This required resequencing some activities, pulling some processes out to be performed at another time, or eliminating inherited tasks from legacy processes that were no longer needed (see "Redesign Work" in Lessons Learned section for details). Redwood was also very helpful during process redesign because their experts helped FSS understand industry best practices.

Just as FSS had done with Engineering Plastics, FSS closely involved the three business groups to verify that the software robots were executing tasks as expected. For this phase, FSS included extra checkpoints in the software to build trust and to gain stakeholder buy-in.

By the end of Phase I, FSS had exceeded its business case by automating 89% of its manual tasks. Quality also improved. de Haas offered the example of booking journal entries. Prior to robotization, journal entries came in on spreadsheets or emails to be processed by humans. After robotization, journal entries were input directly into the software robots. The robots evaluate the entry, post it, and send it back to the business units without human intervention: *"Nobody's touching it. Everything is done automatically through the robot, which of course is good because of speed but also the quality. …Previously we still had discussions like, 'book a thousand Euros,' and somebody else said, 'no, no, it was 10,000 Euros, you didn't hear me well.' Those kind of issues are gone now."* – Theo de Haas

Phase I was completed in November 2015. FSS focused on year-end close before beginning Phase II.

PHASE II: FEBRUARY TO AUGUST 2016

Phase II sought to bring six more business groups onto the RPA platform. In total, 130 country codes from across the business groups were to be migrated to RoboClose®, which added many more users to the RPA program. For this phase, FSS had learned enough about the software to take charge of implementing the business rules. The team built three templates for three of the six business groups, configured the software, and ran user acceptance tests by April 2016. Once again, the robots worked as expected and the three business groups went live the following month. As FSS's RPA team gained more experience, their ability to onboard new companies accelerated: *"We're now in a situation that we can do one whole company code per week. We will set the business rules, we'll test it, and we do a full production in one week which is, if you look from an automation point of view, it's unheard because it's not something you do with an SAP implementation. You don't do an SAP implementation in a week."* – Theo de Haas.

The remaining three business groups were migrated a few months later.

BUSINESS RESULTS FROM PHASES I AND II

"We went globally live for Europe, for China, for APAC, for the US. We didn't have any glitches…it worked like a charm." – Theo de Haas

As highlighted in the Introduction (Table 1), DSM achieved multiple business benefits from automation, including FTE savings and faster delivery of financial close from 15 to three days. As a consequence of automation, fewer humans were needed to perform the month-end close process. In total, about 45 fewer FTEs were needed to complete the process. The human work that remained was shifted away from doing transactions to more value-added monitoring, auditing, and judging the results. Excess labor has been redeployed to other tasks when feasible.

6. WHAT'S NEXT FOR RPA?

FSS has both near-term and long-term plans for automation. In the near-term, DSM's Latin America business units that use Oracle as their ERP system

will be migrated to the RPA platform for financial close. Eliminating some of the extra checkpoints that were put in to build confidence during the startup phase will also further optimize the financial close automation. de Haas said, *"We can start optimizing by taking some controls out… we now trust how it works."*

Thus far, automation has been applied to the tasks performed by the captive center in India, where 80% of the financial close processes take place. But recall that FSS also has regional business centers that perform the other 20% of activities. Might some of these activities be automated as well? Hussain thinks so: *"There are also critical activities that are performed in the business. So we feel that we can bring these automation solutions over there and see if there is something that can be automated. We would then require the involvement of the business and make sure that all the businesses performed the activities similarly."*

Beyond financial close, FSS is considering RPA for accounts payable, accounts receivable, and credit management. As de Haas explained: *"RPA is here to stay so it's not something that will go away. It's not hype, it's not something that will pass in three or four months."*

Next, we discuss the DSM case and compare its lessons with our prior RPA case studies.

7. CASE DISCUSSION AND LESSONS LEARNED

What might other organizations considering RPA learn from DSM? We discuss four lessons pertaining to **project management** – managing the phases of the automation program, and four lessons pertaining to **change management** – managing the stakeholders affected by change including senior executives, business groups, employees working in shared services, and the IT function. We see project management as *"doing the thing right"* and change management as *"doing the right thing"* [Lacity (2008)].

Best practices for project management include (1) letting business operations lead RPA, (2) picking the right automation approach, (3) selecting the right implementation provider, and (4) redesigning processes to maximize the benefits and minimize the risks of automation.

7.1 Let business operations lead RPA

Potential service automation adopters often ask, *"Where is service automation launched – in business operations, IT or in outsourcing provider firms?"* Across our 15 RPA client adoption stories, 13 automation programs were led by business operations groups, including shared services groups, and two were led by IT. RPA's appeal is that the tools are designed to be used by subject matter experts (SMEs) rather than by IT programmers. In fact, as we have noted, it is more accurate to say that RPA users *configure* the software robots rather than *program* the robots. RPA recognizes that it is cheaper, better, and faster to train SMEs to do their own automations rather than have SMEs explain their deep domain understanding to an IT software developer who then explains it to a team of IT coders. Because RPA tools are designed for SMEs, RPA adoptions are primarily initiated and led by business operations. At DSM, de Haas explained why FSS led the project:

> "It's not an IT project, it's a business project with a small IT component. 99% of the project is about business rules; it's about making sure that the processes work so we did it with business process experts."

7.2 Pick the right RPA approach: screen automation versus process automation

During the course of our research we have learned that by mid-2016 over 40 software providers were marketing their tools as RPA. But these providers have very different approaches to automation. Some offer quick and cheap solutions that are deployed on desktops. These tools are suited for organizations that want to democratize the workforce and allow individuals to control the automation of their own work. Other RPA providers, including Redwood, aim to automate enterprise transactions on a platform that is secure, available, and controlled. For DSM, this latter approach fit their needs because they aimed to automate financial close – an extremely important process to control and secure.

de Haas also advised prospective RPA buyers to consider the total cost of ownership, not just the cost of the RPA software license. Screen automation software is typically cheaper and easier to learn than process

automation software, but total costs of automation need to consider the full development and long-term maintenance costs. de Haas explained: *"A lot of RPA vendors are really just doing screen scraping, which requires a lot of maintenance if you want to change it. I think the biggest advantage that we have with Redwood is that everything is controlled by business rules. So my advice to companies who really want to do this is you should do process automation and not screen automation. If you do this at the screen level, probably you'll wind up be having even more problems that erode into your savings because you have huge maintenance on your hands."*

Neil Kinson, Chief of Staff for Redwood, concurred. According to Kinson, companies should be aware that screen automation tools could result in *"a plethora of point solutions dealing with individual micro-process or meta-robots that becomes unmanageable. If you roll out a new version of ERP, suddenly you'll break all of your robots, or at least you have to retrain them. And more importantly, it's very difficult for the IT function to quantify and define and control those changes."*

This is something we have also observed in several case studies examined in our new research since February 2016.

7.3 Pick the right implementation partner

Once organizations pick an RPA approach, they also have to pick an RPA tool and an implementation partner, which may or may not be one and the same. Picking the tool is actually the easy part. A number of advisory firms now have RPA practices to advise clients on technical capabilities and total cost of ownership for the more established RPA tools. The harder part is picking the right implementation partner. The right partner needs real subject matter expertise and enough excess talent to devote FTEs with those rare skills to the client organization for the entire engagement. Implementation partners also need prior experience with the tool and they should be willing to help the client build a mature RPA capability so the client can function independently after the engagement.

If real estate success is all about, *"location, location, location."* then RPA success is all about "subject matter experts, subject matter experts, subject matter experts." The focus on SMEs is also why DSM selected Redwood; Redwood helped FSS tweak its processes based on industry best practices, not just on what FSS was currently doing. Mohammad-Sajjad Hussain – Lead Business Process Expert, DSM Business Services India, explained: *"Redwood has finance experts and they have good knowledge about industry best practices. So when we were explaining our way or working, we were also confronted with their expert outsider view. They asked us why we do these things. So that was a challenge but also a learning for us."*

Sometimes implementation providers oversell, leading to higher costs or project delays. As one famous outsourcing theory argues, providers are incentivized to behave opportunistically, that is to pursue self-interest with guile and to make *"false or empty, that is self-disbelieved, threats and promises."*[7] Research has shown that the risks of provider opportunism can be mediated with strong contractual governance, but rigorous contract negotiations and contract monitoring increase transaction costs.[8]

Another way to mitigate opportunism is to invest in strong relational governance based on mutual obligations, trust, and co-commitment.[9] DSM credits the good relationship with Redwood, in part, because of its prior relationship, but also to the fact that Redwood's RPA sales team was also part of the RPA delivery team. This ensured that the sales people did not over-sell, over-promise, or over-commit. DSM interviewees also praised Redwood's subject matter expertise and cooperation. Hussain said: *"The engagement with them was excellent. They had very good expert knowledge and they were very patient and we all worked well together."*

This endorsement is particularly impressive given that the Redwood team was remotely located. Hussain continued: *"We weren't all sitting in a room face-to-face, but we worked well in a virtual environment. Amongst us, we have the understanding that if we see any issues or we see something that's not going well, that we communicate to each other. We partnered with them very well, even though it was a virtual team, we never felt that we are distant from them or that we don't understand each other."*

Next, we address **change management,** and how best to address the concerns of senior executives and business unit leaders, the employees working in shared services, and the IT function.

[7] The idea of vendor opportunism comes from Transaction Cost Economics, a theory about make or buy decisions that assumes providers will take advantage of customers if given the opportunity to do so [Williamson (1975)]

[8] For example, Williamson (1979, 1991) argues that contractual governance can mitigate the risks of vendor opportunism.

[9] For a review of the empirical research on relational governance, see: Lacity et al. (2016).

7.5 Show RPA's capabilities to senior managers and business units

Decades of research identify senior management support as a critical factor for project success.[10] Automation programs are no different – the client organizations in our study achieved, with C-suite support, the most strategic benefits from service automation. At DSM, the Director of FSS was quick to gain the support of the CFO during the PoC phase. The CFO's major concern was that any automation needed to have built-in audit trails to show exactly what the software robots were doing at each step in the process. Similarly, the business units also wanted confirmation of compliance by any work being done by software robots.

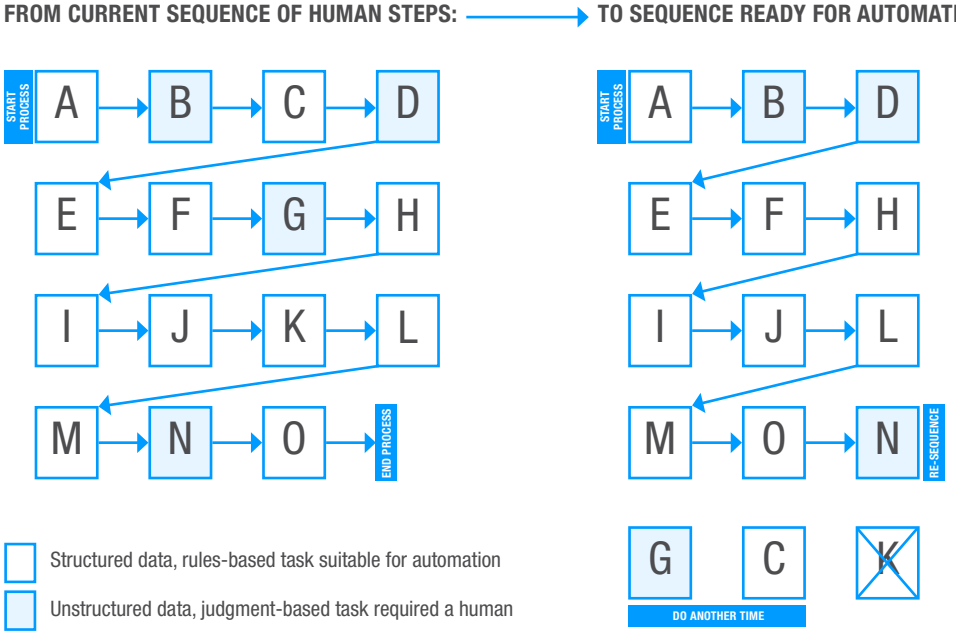
While the RPA program team quickly became convinced that the robots would not go rogue, senior managers and business unit managers needed more sustained evidence before trusting the software robots. de Haas explained: *"Because trust is one of the key things you need when you do RPA, it cannot be a black box. It's really necessary for people to build the trust that the robot is not making mistakes or screwing up figures."*

Hussain added that building up this trust delayed the project a bit, but that this aided user acceptance. He said: *"People had to test them and certify that they are comfortable that the robot is doing exactly what they would do manually. So, it took some weeks for them to understand what the robot is doing."*

7.6 Prepare retained employees

Like all our RPA case studies, FSS had to define what the new organization would look like after robotization. The immediate task was to define which tasks software robots would do and which tasks the remaining employees would do. Furthermore, the employees needed to be trained on how to work with robots. In general, the employees working with the software robots welcomed automation. Hussain said: *"It's about working with the new tools so that it's something*

Figure 1: Optimizing the sequence of steps in an end-to-end process



Note: The left-hand side of the figure depicts an end-to-end process with 15 steps currently being done by a human. Some of those steps, depicted with clear boxes, entail structured data and rules based processes, such as looking up data fields in an existing system of record. Some tasks, depicted with blue boxes, require judgment, interpretation, or problem-solving skills. When examining this process for automation, only the clear boxes are suitable for RPA, but the current sequencing would require the human to intervene four times. To optimize the sequence for automation, an RPA team might realize that some tasks are not needed and can be eliminated (like task K depicted on the right hand side), some tasks can be pulled out of the process and done at another time (like tasks G and C), and some tasks that require human intervention might be batched (like B and D) or re-sequenced (like N).

[10] These references show years of research tying senior management support to project success: Standish Group (2015), Sabherwal et al. (2006), Lacity (2008), Nelson (2007).

exciting for people who are just used to doing the activities manually."

Neil Kinson, Chief of Staff for Redwood, added: *"So RPA is not just about taking FTEs out, it was about raising the level from pure transactional to more rewarding work, creating a higher quality of work."*

7.7 Ease transition for redundant employees

Across our RPA cases, the topic of redundancy is always sensitive. Organizations like to share the stories of upgrading the skills of the retained employees or taking on more work without adding more headcount. The reality is that as RPA scales, many companies need fewer employees. Companies need to develop plans for redundant employees. Across our cases, organizations either waited for natural attrition to gradually ratchet down headcount, offered early retirement, or offered career counseling for redundant employees that could not be deployed to other jobs within the organization.

At DSM, the majority of people were reassigned to higher value tasks, such as reporting and redesigning processes. de Haas explained: *"People want to do that [take on more challenging work] but also of course, we assess them to also make sure that they really can make that step. We also give them the chance of doing it."*

Some redundant people, mostly middle managers, were offered professional career counseling and used the opportunity to pursue lifelong dreams (for example, in one case, helping orphans). The long-term challenge is to figure out the career paths in shared services for the next decade. This is so important that we have devoted an entire section on the topic (see RPA Challenges below).

7.8 Bring IT on board early

Across our 15 RPA cases, business operations questioned when or if to bring in the IT department. Some RPA champions in our other case studies initially excluded IT at the onset for two reasons: (1) service automation was seen as a business operations program since it required process and subject matter expertise, not IT programming skills, and (2) fears that IT would beleaguer the adoption with bureaucracy. In most such instances, however, clients found, in hindsight, that IT has an important role to play. Clients learned the importance of involving the IT department from the beginning so that IT can help validate the RPA software as enterprise-worthy, manage how software robots access existing systems, and manage the infrastructure so that it is available, secure, and scalable [Overby (2016)].

At DSM, FSS informed the Global IT Leadership Team after the robots were in production during May of 2016. FSS gave the IT leaders a demonstration of the product, explained how it worked, and showed them the business results in terms of cost savings and quality improvement. According to de Haas, the CIO was very

impressed with the speed of project delivery and the results. Why didn't FSS bring IT in the loop earlier? As noted above, FSS considers robotization to be a business project, not an IT project. But eventually IT needs to be informed and needs to help manage the production software.

Redwood took the lead for educating DSM's IT department because it was in a better position to explain the technical requirements of the software than FSS. Edwin Klijsen, Director of Financial Transformation for Redwood, explained the role of IT in RPA: *"The business user will drive the solution but at a certain point, of course, IT also has roles within this; it's an IT solution in the end, so there needs to be maintenance and support and input also from IT side."*

Klijsen also concurred that IT needs be brought in early in the process: *"When considering enterprise grade robotic solutions then collaboration with IT is a must. Any enterprise class robotic system needs to meet the IT security and governance requirements of the organization."*

8. FUTURE RPA CHALLENGES: CASE REFLECTIONS

In addition to the best practices emerging from our research, there are also "future of work" challenges with which all companies are grappling. In the long-term, what career paths do companies offer humans after robotization so that organizations retain enough knowledge and keep the humans engaged? What will happen to employees working in low cost countries? What is the corporate responsibility and duty to them? There are no easy answers, but we discuss some of these issues below.

8.1 Career paths in shared services after robotization

What will the career paths for internal employees look like in a world full of software robots? All the companies in our study asked this question, and certainly DSM is concerned about the future skills it will need in a highly robotized environment. We note that this question has been raised many times before in the context of outsourcing. In our outsourcing research on hundreds of companies collected over two decades, companies asked, *"What will the career paths for our internal employees look like in a world full of outsourced labor?"* We think the answer to both questions is the same.

Transitioning to software robots (or to BPO providers) can be tough on workers in the retained organization if steps are not taken to help them succeed in the new environment. Their roles will often shift and they will find themselves charged with managing and coordinating the work done by others rather than executing tasks. But the real aim should be focusing and empowering staff on customer service and business enablement. Too many times the employees are bogged down in the drudgery of transaction execution rather than exploiting the data collected to improve business operations. Kinson explained: "*It's a classic positioning model but it continues to be true that very few finance functions have the time and capacity to provide the insight to the business that the numbers are telling them. Their entire capacity and effort is making sure their auditors are happy and the numbers are produced in the time available and that they're accurate enough so that the CFO doesn't get in trouble."*

Automation can help free up employees for more valuable tasks, the consequence of which will be a new organizational design for shared services; a design shaped like a *"diamond"* instead of a *"pyramid"* (Figure 2) [Lacity et al. (2014)].

Pyramids are heavily populated with employees, most of whom are at the bottom of the pyramid doing transactional work. The benefit of this design is that employees continually build valuable, organization-specific experience as they are promoted higher up the pyramid. The pyramid model is strong on retained knowledge, but it is also costly. Shared services managers trying to recruit college graduates must compete with RPA or BPO providers who can court them with far richer career paths and many more peers. The model also tends to rely on staff augmentation with expensive domestic workers to fill in skills gaps, and to scale up resources. A significant class of middle managers who manage both employees and supplemental staff also characterizes the pyramid model.[11] Diamond-shaped retained organizations replace the heavy bottom of the pyramid with robots. Machines perform many transactional activities that were once performed by employees now. There are fewer middle managers needed, a view also predicted by de Haas: *"Robotization will affect the middle layer. I think a lot of those activities for instance, in our case it's putting in an invoice into a system or sending out letters to a supplier saying, 'your invoice needs to be*

[11] We first identified the shift from pyramids to diamonds in Chapter Six of Lacity and Willcocks (2015b).

resubmitted.' Robots will do all that. People won't do that anymore so they have to reskill themselves to a higher level and we also have to face that a lot of those people in the middle layer are not able to do that or have no appetite of doing that."

But one statistic from our RPA research foretells a possible future with fewer middle managers supervising lower level employees: in one U.K. utility, only two human beings were needed to manage 300 software robots that were performing the work of 600 FTEs [Willcocks and Lacity (2016]! Our own view is that the type and number of tasks in the middle layer will increase, not least because the cognitive automation that many see replacing tasks in the middle layer may be uncommercial, technically unready, and more difficult to get up and running, at least over the next three years, than many are predicting.

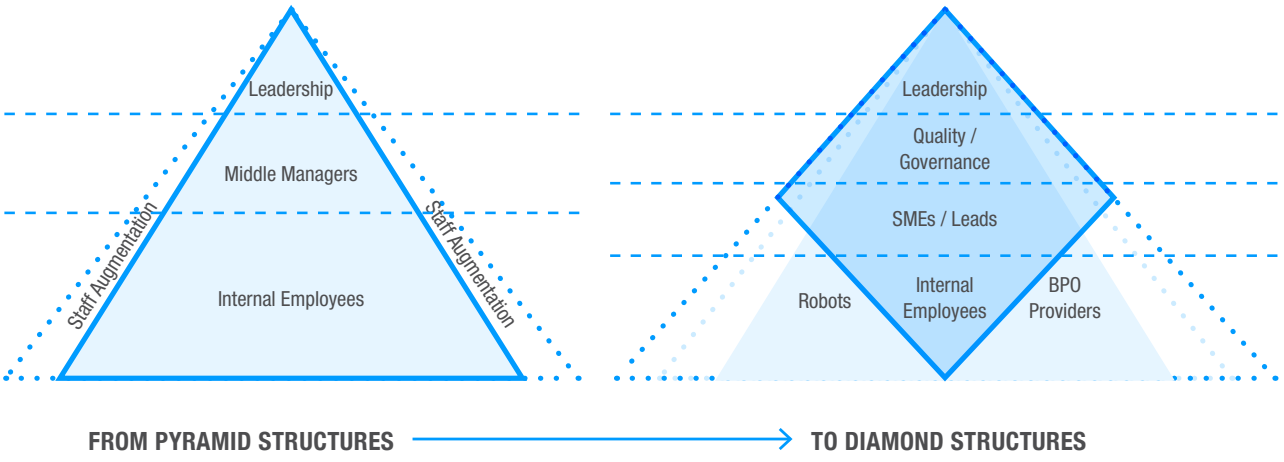
The diamond-shaped organization needs more subject matter experts, quality assurance, and governance skills to coordinate services with internal business units and with RPA and BPO providers. de Haas predicts that the humans who have these skills will be transient, with rich careers across organizations rather than within a single organization. He said: *"Highly educated people that are extremely flexible, that have a change mentality will work in many companies. We won't get people that work 25 and 30 years for a company. I believe that they will work on projects and when the project is complete, then there's nothing of interest, so they move onto the next company."*

The benefits of the diamond-shaped retained organization are lower costs, access to providers with best-of-breed skills, and greater flexibility because robots can more easily adapt to increases or decreases in service volumes.

8.2 Impact on labor in India and other low cost locations

Another major concern people share is the potential impact of automation on jobs in low cost countries like India. Initially, India's vibrant ITO and BPO sector grew based on the value of lower cost labor. In a highly automated world, labor arbitrage is no longer a compelling value proposition. Among our 15 RPA case studies, four had operations in India. DSM and Xchanging had captive centers in India and both companies chose to house their RPA programs in the Indian captive centers. de Haas, as previously stated, estimated that 99% of an RPA project entails defining business rules, which requires deep subject matter expertise. This is why FSS chose to implement RPA in the Indian delivery center; Hyderabad is where the most of the SMEs on financial close are located. Locating the software robots near the SMEs seems optimal because processes need both human expertise and robots to function as a team to be most effective.

Figure 2: Transforming shared services from pyramids to diamonds



Note: The figure of the pyramid and diamond was adapted with permission from Jim Lammers of Express Scripts and from Sandy Ogg of Unilver.

Two of our cases, a European utility and a U.K.-based telecommunication company, had outsourced to BPO providers in India. They were among our earliest RPA adopters back in 2008. Both initially tried to partner with their BPO providers to do or to help with automation, but the FTE-based contracts, the BPO provider business models, and perhaps other undisclosed reasons resulted in the clients taking back the processes. Both these companies chose to reshore processes after RPA. The implication seems clear enough: Indian-based providers need to adapt, a message often conveyed in the Indian media.

We offer two insights here. First, the Indian-based providers are also innovating, which will increase their value propositions and bring new opportunities for growth. Indian-based BPO and ITO providers like Infosys, Wipro, TCS, and Tech Mahindra have all introduced automation to produce FTE savings. Hence, although fewer people will be needed for low-level, structured, and rule-based tasks, more jobs could be created for mid-level, unstructured, and creative tasks. According to Som Mittal, president of NASSCOM: *"Growth in future will be driven by new services/solutions and not more of the same. The industry has started to make significant investments in tools, technology and talent to build appropriate solutions and communicate the value proposition"* [Phadnis (2013)].

Second, many young and educated Indian professionals do not want the boring, repetitive jobs that Western-based organizations send offshore. In 2008, we published the first study that examined the reasons for the high turnover rates in Indian-based BPO and ITO providers, which were reported at the time to be as high as 80% in the IT services sector and as high as 100% for Indian call centers [Gupta (2001, Mitchell (2004, 2005)]. Based on interviews, we found that turnover was significantly related to task variety and complexity [Lacity et al. (2008), Iyer, V. (2011)]. A strong theme throughout the interviews was that Indian professionals want challenging jobs, just like their U.S./Western counterparts. The professionals who were most dissatisfied with their Jobs were mostly upset about the lack of task variety and low skill set utilization. For example, one participant complained: *"I have been put into testing and coding and now it is kind of maintenance phase. Now I am not able to use my skill set much. I am not satisfied with the kind of work I am doing. Every alternate day I go to my manager and I tell him that I am not satisfied with the kind of work I am getting and I need more challenging work so I can improve my skills."* Another interviewee complained, *"I just fix errors.….I utilize only 20% of my knowledge."* Kinson also corroborated that Indian professionals want interesting work: *"India has significant attrition because Indian employees earn multiple degrees, they've earned MBAs, and they're just not satisfied doing low level work."*

9. CONCLUSION

RPA really took off during 2016 and 2017, with an estimated 100% annual growth rate for the 2006-18 period [Burnett (2016)]. The Royal DSM–Redwood case demonstrates the challenges, implementation lessons, and multiple organizational benefits possible from this rising phenomenon. It also demonstrates how, like several other service automation tools available, Redwood robotics can be scaled to enterprise level. Indeed, the second half of 2016 found DSM extending robotization well beyond its initial investment into financial close processes and its Indian shared service center. And beyond financial close, FSS was looking to robotize accounts payable, accounts receivable, and credit management.

The DSM case makes clear, however, that when it comes to achieving business benefits good management, amplified by new technology, really does make the difference. We identified eight action principles in the case: let business operations lead RPA; pick the right RPA tool; select the right partner; redesign work; show RPA tools to senior managers and business units; prepare retained employees; ease transition for redundant employees; and bring IT on board early. At DSM, these formed the strong foundation for growing and scaling service robotization in the enterprise globally. These findings gel very well with our findings in previous shared service cases.The Royal DSM case also demonstrates that RPA is best treated as a strategic long-term investment and not as a one-off tactical initiative. Strategically, this brings to the fore several challenges that we, at the time, emphasized less in earlier cases. As we discussed above, strategic use of RPA requires careful thought, and much preemptive focus on future work design, change management, and the skills implications and human-machine balance for the emerging workforce. Our wider evidence so far, documented in our recent book and many papers [Willcocks and Lacity (2016), Willcocks (2016), Lacity and Willcocks (2016c)], is that service automation, applied strategically to scale across multiple processes, will reshape how work is achieved, its location, and the human skills mix needed, against a context of ever rising information workloads for organizations. These were certainly the major issues Royal DSM and Redwood were concerning themselves with, as they considered future robotization in the Royal DSM global financial shared service operations.

References

Burnett, S. 2016, "Brexit, automation and the impact on outsourcing," presentation at the NOA Breakfast meeting "Outsourcing in the Brexit climate," August 3

de Haas, T., 2016, "Robotization in shared services," presentation at First European Innovation Summit, Weisbaden, 27-28 June

Gupta, P., 2001, "Growth scenario of IT industries in India," Communications of the ACM 44:7, 40–41

Iyer, V., 2011, "Understanding turnover intentions and behavior of Indian IS professionals: a qualitative study of organizational justice, job satisfaction and social norms," University of Missouri-Saint Louis

Lacity, M., 2008, "Editor's Introduction," in Willcocks, L., and A. Lee, Major currents in information systems: the management of information systems, Volume 4, Sage

Lacity, M., V. Iyer, and P. Rudramuniyaiah, 2008, "Turnover intentions of Indian IS professionals," Information Systems Frontiers, Special Issue on Outsourcing of IT Services, 10, 2: 225-241

Lacity, M., and L. Willcocks, 2015a, "What knowledge workers stand to gain from automation," Harvard Business Review Online, June 19

Lacity, M., and L. Willcocks, 2015b, Nine keys to world-class business process outsourcing, Bloomsbury Publishing

Lacity, M., and L. Willcocks, 2016a, "Will your next relationship be with a robot? A look at how service automation may impact outsourcing," Pulse Magazine, March/April, 24-27

Lacity, M., and L. Willcocks, 2016b, "Clients slower than providers to embrace service automation," Pulse Magazine, May/June

Lacity, M., and L. Willcocks, 2016c, Robotic Process Automation at Telefonica O2," MISQ Executive 15:1, 21-36

Lacity, M., S. Khan, and A. Yan, 2016, "Review of the empirical business services sourcing literature: an update and future directions," Journal of Information Technology 31:2, 1-60

Lacity, M., P. Willcocks, and A. Burgess, 2014, The rise of legal services outsourcing, Bloomsbury

Willcocks, L., and M. Lacity, 2015a, "Businesses will increasingly use robots to deal with the explosion of data," LSE Business Review Online, September 15

Willcocks, L., and M. Lacity, 2015b, "Will robots replace humans?" LSE Business Review Online, October 6

Willcocks, L., and M. Lacity, 2015c, "Poppy appeal: RPA at BPO service providers," Professional Outsourcing Magazine, Winter issue, 12-18

Willcocks, L., and M. Lacity, 2015d, "IT, phone home: Leslie Willcocks and Mary Lacity on Telefonica O2's RPA," Professional Outsourcing Magazine, Autumn issue, 26-34

Willcocks, L., and M. Lacity, 2016, Service automation: robots and the future of work, SB Publishing

Lacity, M., L. Willcocks, and A. Craig, 2015, "Robotic Process Automation at Telefónica O2," The Outsourcing Unit Working Research Paper Series

Lacity, M., L. Willcocks, and A. Yan, 2015, "Are the robots really coming? Service Automation Survey findings," Pulse Magazine 17, 14-21

Mitchell, A., 2004, "Offshore labor markets impact IT outsourcing," Technewsworld, September 28, http://bit.ly/2veBWAB

Mitchell, A., 2005, "India maintains outsourcing advantage," Ecommerce Times, May 3, http://bit.ly/1Y6nWAv

Nelson, R., 2007, "IT project management: infamous failure, classic mistakes, and best practices," MIS Quarterly Executive 6:2, 67-78

Overby, S., 2016, "Should CIO be chief robot wranglers," CIO online, July 26, http://bit.ly/29YeqyH

Phadnis, S., 2013, "IT to reskill staff as technology reshapes business," The Times of India, September 18, http://bit.ly/2vXMHFE

Sabherwal, R., A. Jeyaraj, and C. Chowa, 2006, "Information system success: individual and organizational determinants," Management Science 52:12, 1849-1864

Standish Group, 2015, "Standish Group 2015 Chaos Report - Q&A with Jennifer Lynch," http://bit.ly/1Y6nWAv

Willcocks, L., 2016, "How organisations can embrace automation," European Business Review, April, http://bit.ly/1CJZJ5f

Willcocks, L., M. Lacity, and A. Craig, 2015, "Robotic Process Automation at Xchanging," The Outsourcing Unit Working Research Paper Series

Williamson, O., 1975, Markets and hierarchies: analysis and antitrust implications, Free Press

Williamson, O., 1979, "Transaction cost economics: the governance of contractual relations," Journal of Law and Economics 22, 233–261

Williamson, O., 1991, "Comparative economic organization: the analysis of discrete structural alternatives," Administrative Science Quarterly 36:2, 269–296

The financial auditing of distributed ledgers, blockchain, and cryptocurrencies

DANIEL BROBY | Director, Centre for Financial Regulation and Innovation, Strathclyde Business School
GREIG PAUL | Cyber-security Research Engineer, University of Strathclyde

ABSTRACT

The internet and digital transfer of money is set to fundamentally change the way financial audits are conducted. This paper critically assesses the way that such assets are currently audited when stored in distributed ledgers, transmitted via a blockchain, or whose value is stored in crypto rather than sovereign currency form. We identify the self-verifying nature of such financial data that negates the need for traditional audit methods. Despite the promise of such methods, we highlight the many weaknesses that still exist in blockchain technologies and how these present issues for verification. We address distributed transaction and custody records and how these present auditing challenges. Finally, we suggest how auditors can use smart contracts to address these issues and at the same time provide arbitration and oversight. Our contribution is to propose a protocol to audit the movement of blockchain transmitted funds so as to make them more robust going forward.

1. INTRODUCTION

An audit is an official examination and verification of financial accounts and records [Whittington and Pany (2012)]. It can be conducted either internally and/or externally by a qualified third party. The principles of modern auditing, as first laid out in Brink (1988), revolve around a statement of responsibilities, a common body of knowledge and standards alongside a code of conduct. These collectively encompass the pre- and post-examinations of a corporation's financial revenues and disbursements, and a review of its soundness, effectiveness, and compliance with both internal and external controls. We argue that the application of these in a corporate setting needs to adjust and evolve to take into account the distributed nature of financial information stored on distributed ledgers, blockchains, and/or in cryptocurrencies. All the current norms are being challenged by the advent of these three new modes of digital asset storage and transmission. This paper investigates these phenomena and addresses the problem of how financial audits have to adapt to reflect them.

IFAC (2009), which encapsulates the international standards of auditing, was devised by financial practitioners, not experts in distributed technology and software protocols. And, despite Francis' (2004) scholarly view that auditing is inexpensive, informative, and positively associated with earnings quality, but impacted by the legislative framework, audit risks do exist, and tend to become amplified when technological complexity is taken into account.

In order to understand the auditing challenges, we offer a brief explanation of blockchain [Nakamoto (2008)]. Each block in a blockchain may contain one or more transactions, with the block header referencing the contents of the previous block in the chain. This ensures that the content of a block cannot be tampered with after its creation, without other parties being able to detect and reject this manipulation. The chain, therefore, acts as a distributed ledger, where each party holds and validates it on an ongoing basis. Indeed, the processing of each transaction is, to some extent, an audit in itself, since every participant in the network ensures that all credits are the result of permitted debits. As a result, Rezaee and Reinstein (1998) argue that electronic data and the Internet "signal the end of the traditional audit." That said, the need for corporate audits for financial purposes is self-evident and we argue that it is just the nature of the audit that must change.

2. TRANSACTION MALLEABILITY

The primary function of financial reporting is the recognition of revenues and expenses, safeguarding of cash, and control over procurements [Rogers et al. (2004)]. A major challenge to this process, and hence for auditors in a blockchain world, is that of transaction malleability. This is where a transaction can be changed after it has occurred. Andrychowicz et al. (2015) showed that the issue arises due to the implementation of the transaction ID algorithm within bitcoin. Malleability makes it possible for a party relaying a transaction (such as a miner or other relay) to modify the transaction in a trivial manner, such that the contents of the transaction remains materially unchanged (with the transaction signature remaining valid). The transaction ID (which is a hash of the transaction data itself) is altered to differ from that originally produced by the party generating the transaction.[1]

The malleability of bitcoin transactions can have two potential implications for auditing a blockchain. Firstly, malleability makes it possible for a transaction to be generated under one ID, yet broadcast and incorporated into the blockchain under another transaction ID. This naturally presents a challenge for auditors, since typically a transaction ID would be considered as a unique identifier. If malleable blockchain payments were frequent occurrences, the reconciliation of payment authorizations from sender against blockchain entries may be difficult.

As a consequence of the above, there is potential for double-payment fraud; something that auditors have to be vigilant about. For example, a participant in the blockchain, particularly one using simple payment verification (SPV) rather than downloading and monitoring the full blockchain, could be tricked into issuing payment twice, with a party claiming the payment did not go through, showing the lack of existence of a transaction under the ID generated by the sender. If the sender does not verify their previous transactions properly, checking the blockchain for all recent transactions, they may not see the transaction appear under an additional transaction ID, resulting in a double payment being made. Accounting for such double payments in an audit may be a challenge, particularly where auditors are not familiar with the technical constraints and restrictions in the implementational quirks of blockchains, such as bitcoin in this case.

[1] http://bit.ly/2ePXiOE

3. DAO TYPE ISSUES

The world of digital money not only covers transmission and storage but also smart contracts, hence the concept of a DAO (digital autonomous organization) has been floated. A DAO is designed to resemble in many ways a conventional corporation, with its own rules and regulations, although it does not inherently exist as a legal person within any given jurisdiction [Ringelstein and Staab (2009)]. This clearly presents an issue for an audit that is focused on a legal entity. The original DAO within Ethereum was built as a form of organization, whereby those who "bought into" the DAO became stakeholders. Those holding tokens issued from the original sale were then viewed as shareholders, able to vote on different kinds of proposal. The rules of the organization (themselves able to be altered through a voting process) would then be used to vote on proposals regarding how the organization's funds are to be spent. In essence, a DAO presents a form of cryptographically enforceable articles of association; DAO-controlled funds cannot be spent without the cryptographic agreement of stakeholders, per the rules defined and voted on by stakeholders.

Various audit challenges are posed by DAO-type structures, not least that of jurisdiction of the entity, and how judgments could be enforced against it. Since the DAO in itself is not a legal entity, its position in law is unclear. In addition, were a judgment to be issued against a DAO, the means of enforcement against it would also be unclear; without agreement of a majority of shareholders, or whatever is defined in the DAO's smart contract rules, it would not be possible for funds to be taken from the organization. Consequently, we recommend that assets held within a DAO should be carefully considered, in particular around the requirements needed to be satisfied so that they can be accessed or spent.

4. LONG-TERM BLOCKCHAIN FORKS

Another challenge to the soundness of an audit is the potential for long term blockchain forks [Gervais et al. (2016)]. A fork is formed when a blockchain has two potential paths forward, either with regard to its transaction history or a new rule. While transient blockchain forks are a fairly regular occurrence, where more than one valid block is produced as the next block in close time proximity, there is another scenario, potentially of concern to auditors. In the event of a blockchain (itself inherently decentralized with no one party in charge) experiencing a breakdown in relations within portions of the community, a long-term fork is a potential outcome. In this scenario, two or more distinct groups would only recognize their own version of the blockchain as the correct chain, and refuse to recognize the others. This would typically occur as a result of network enforcement of rules. Examples of this may include alterations to validity requirements on transactions, or of blocks. For example, the bitcoin maximum block size is one megabyte, and raising this would require a fork to the blockchain, since larger blocks would be viewed as invalid by those following the older rules.

> "When recording the balance of accounts holding cryptocurrencies or other such commodities, one accounting challenge faced by auditors is that of ascertaining the currency in which the audit should report the overall balance of funds."

Auditors need to be cognizant of situations where a community formed around the concept of larger blocks at a given raised limit (say two megabytes). In such a scenario, one group of miners may decide to mine and produce larger blocks, while others reject these blocks and continue to produce their own blocks with a maximum size of one megabyte. At this point, a divergence would occur. Transactions taking place prior to the fork would be present on both chains. Transactions taking place after the fork may appear on one, or both, chains. To further complicate matters, blocks mined on one chain may also be valid on another, depending on the nature of the fork. For example, in the scenario of a block size increase, blocks mined while adhering to the 1 MB size limit would presumably also be valid on the fork permitting larger blocks, provided they were mined with the correct parent block header hash, thus advancing the chain correctly.

5. SHORT-TERM BLOCKCHAIN FORKS

Short-term blockchain forks are a somewhat more regular occurrence. As a result, they present auditors with more frequent issues. In bitcoin, this happens in the period between blocks being produced (the mean inter-block period is regularly recalibrated with block difficulty adjustments to be 10 minutes). Where two

miners near-simultaneously discover a valid solution for the next block, one block will become the successor block, and the other will become an orphan block. The block that is propagated to the majority of nodes first will most likely become the valid successor, since they will attempt to build upon that block, and more parties attempting to mine upon it means that this block is most likely to have a successor. Once one side of the chain becomes longer, one block will orphan, with its transactions returned to the pool of pending transactions, and the block recognized as invalid, due to a longer chain existing without incorporating that block.

The risk of short-term forks, referred to as orphan blocks, is minimal, since it occurs regularly in the bitcoin blockchain (around once per day is not uncommon), and participants can handle the scenario elegantly. For an auditor, however, the potential for orphan blocks makes it important to ensure that the audit only covers blocks that have sufficient proof of work upon them to make any future re-arrangement orphaning those blocks infeasible. One significant factor to note is that bitcoin will accept any longer chain at any point in future, if such a chain exists. There is, therefore, no time period beyond which it can be guaranteed that no alternative longer chain will emerge. At any time, a longer chain being announced to the network would result in the adoption of the longer chain. While past transactions could then be re-broadcast to the network for inclusion, since they were already signed, this introduces the potential for double-spends to occur, where the (previously hidden) chain incorporated a transaction to spend funds that were spent in the (broadcast) chain. This would result in the recipient of the broadcast transaction to lose the received funds in the subsequent reorganization to accept the longer (previously hidden) chain.

6. FINANCIAL CUSTODY

Custody and distributed ledgers need to be audited. Traditional audits inspect the custodial assets held by a legal entity. The role of custodians in the context of distributed ledgers will clearly evolve and as such presents auditors with new challenges. As it currently stands, market infrastructure currently relies on a hierarchy of custodians. A number of legal issues arise from such intermediation. Neoclassical economic theory suggests that we do not know enough about this infrastructure. Financial intermediation chains have contractual ring-fencing from the responsibility of the sub-custodians in this hierarchy. There is, in effect, a behavioral problem at the investor level because of the different bargaining power between the institutional and the public markets. The explanation for this is that the public investors are time poor, have a bias against long term risk, have tax issues, and have a tendency to believe that the future is like the past.

Blockchain technology provides the ability for money to be disintermediated and connected to a central asset ledger via the Internet. Current investors in the public market, who would most benefit from this, do not have the bargaining power to fund such developments. As a result, institutions still have the upper hand. There is a role for auditors in this respect. We need to recognize that even such things as cryptocurrencies involve intermediation. Where the cryptography is provided centrally, the wallet holder effectively becomes the intermediary.

The role of a central third party is not to just keep a ledger, but to ensure they are valid. An auditor has to verify this. In other words, are the distributed ledgers reliable and how do they link to reality? Blockchain explorers can be adapted to provide tools to make it easier to achieve this. Current custody platforms, such as Euroclear, can clearly improve by adopting and adapting their technology but would be at risk of undermining their current business models.

In addition to custody and ownership, auditing is also required to ensure the timestamping of the blockchain, its validity, and its robustness. In the distributed world, there are in fact multiple blockchains, not a single immutable record as the public perceives. As such, a traditional audit of a false fork only provides a detailed record of the records. We return to the latter. In the case of closed, permissioned blockchains, what is required is an audit of who gives permission to the permissioned blockchain. In other words, the audit process should focus on the creation of a chain, not simply give insights into a snapshot in time. At present, reconciliation only occurs at the individual custodian level.

7. CHALLENGES FOR AUDIT

There are many challenges in auditing financial data within a blockchain. One of these is accounting year ends. These are reported at a static point in time. In a blockchain, however, the most recent transactions cannot be guaranteed to be irreversible at a given point in time; their irreversibility is a property of the quantity of mining work carried out on top of those transactions. Each subsequent block mined beyond a given block is referred to as a “confirmation,” signifying that other

miners have agreed that this block is valid, following the necessary rules, and containing only validly signed transactions. We highlight other more technological issues next.

8. MULTI-LOCATION AUDIT RISK

The internet is cross-jurisdictional. This audit issue is addressed by Statement of Auditing Standards (SAS) No 107, which states that an auditor facing such jurisdiction issues has to take into account the nature of the assets and transactions, the centralization of records, the effectiveness of the control environment, the frequency of monitoring, and the materiality of location. That said, the auditing standards incorporate digital storage of value when they were first drafted.

The issue of multi-location was highlighted in July 2017 when a French court gave Alphabet Inc. (Google's parent company) a reprieve from a 1.11bn-euro ($1.27bn) tax bill. The Paris administrative court noted that its subsidiary, Google Ireland Limited, did not have a "permanent establishment" in France. The audit trail, in this instance, being critical in determining jurisdiction.

The need for better auditing standards for digital assets is a fairly new issue. There are a lot of participants in the distributed ledger ecosystem who want credibility and a lot who want reassurance. Clearly, some things are easier to audit than others. The auditing industry needs to define the level of that reassurance. If you go into any form of distributed ledger environment, the cost of audit and regulation currently outweigh the development costs. The current ledger audits are done by the data departments of accounting firms, there being no dedicated audit function that oversees the technological aspect of financial audits.

Current audit practice revolves around accountants entering an organization as external auditors, and carrying out a process of verification of the accounts. With the rise of blockchain, and the potential for non-trivial quantities of assets to be held within, or transferred through, a blockchain, auditors will increasingly find it difficult to ignore these ledgers. The blockchain gives rise to a distributed set of ledgers that bring with them the sort of multi-location audit risks identified by Allen et al. (1998) and Hegazy and Nahass (2012).

Auditing permissioned ledgers involves interrogation of the system. The technology can be audited in real time, but auditing requires an understanding of the context. When you look at a distributed ledger from the perspective of ownership, the coding of a transaction might not be as aligned to the underlying ownership as it exists in the physical world. In a digital context, ownership can also be broken down into describing ownership, protecting ownership, storing ownership, preparing ledgers, the addition of transactions to a ledger, and deciding which ledgers are deemed true and accurate.

9. ISSUES WITH SELF-VERIFICATION

While the design properties of a blockchain being immutable and self-verifying are beneficial to audit, the robustness and reality need to be explored by the auditor. In this respect, Buyya et al. (2008) illustrated how blockchains can be used with cryptographic hashes within decentralized networks. Transactions on a bitcoin-like blockchain are inherently self-verifying. Each transaction is digitally signed, to prove its authenticity, and based upon the outputs of a previous transaction. A transaction can, therefore, be checked by any interested party with access to the blockchain, to ensure that the signature on it is valid, and that it only spends available and unspent funds, satisfying the requirements of the ledger rules.

For example, if party A transfers an asset to party B over a blockchain using this model, a transaction record will be created, whereby party A takes one or more received transactions that they have not yet spent, and specifies party B as the recipient. Any surplus funds can be returned back to party A. The resulting transaction must then be signed by the private key corresponding to each incoming transaction that is used within the transaction. Any party with access to a public key is able to verify if a signature was issued by the corresponding private key holder for that address.

We argue that it is desirable to audit only transactions contained within blocks with a number of confirmations. This indicates when the likelihood of reversal is minimal due to a fork having emerged in the blockchain. It is difficult to quantify the number of confirmations necessary. That said, we suggest that six confirmations is usually sufficient for most large transactions, which would correspond to around a 60 minute delay after a transaction was featured in a block. Despite this, in times of adverse conditions on the blockchain, such as large numbers of mining nodes not properly validating blocks, users have been advised to wait for considerably higher numbers of confirmations. In one case, this was as high as 36 confirmations, reflecting a 6-hour delay.[2]

10. ABILITY TO TRANSACT SILENTLY

Audit helps to detect fraud. Within blockchain-based crypto-currencies, it is possible for parties to create transactions silently, as well as to generate them from any location where the appropriate keys are accessible. Consequently, if a malicious party were to gain access to the private keys for a bitcoin or other wallet, they would be able to generate validly signed transactions from that address at any point in the future, without being located physically within the organization in question. The transactions could be broadcast from any node connected to the bitcoin network, as there is no such concept of authorized signatory, beyond that of anyone holding the correct cryptographic keys. With multi-signature wallets, such as those discussed below as a form of contract, if a party can satisfy the requirements of any given incoming funds, a transaction can be generated from anywhere. Auditors have to find ways to address this issue when ensuring transactions are valid.

In contrast, a regular bank account may require transactions to be initiated from a particular terminal, or have certain approved signatories physically present themselves at the bank to sign a large transaction. Within blockchain, possession of the necessary private keys, or knowledge of the appropriate hashlock condition, is all that is required to perform a transaction from anywhere.

11. ABILITY TO HIDE TRANSACTIONS

For an audit to be effective, it must be bounded to cover a finite period of time, from a starting point to an ending point. The audit should begin at the end of the previous audit, to ensure that transactions do not fall between audits. Within a blockchain, time becomes discrete, rather than continuous, making this process slightly easier. The mean inter-block generation time becomes the increment of time in the chain.

Transactions are not themselves individually timestamped however, so the presence of a transaction within one block does not guarantee that was when the transaction was produced and broadcast. This presents issues for an audit. A time stamp may have been included in an orphan block and now is being included in a new (valid) block. Alternatively, the transaction may have been generated in the past, and then broadcast at a later date. This makes the audit process more complex, particularly if auditing internal controls and procedures needed to initiate transactions, since preauthorized transactions could be broadcast at any later time, thus transmitting the funds long after the authorization was granted.

The timestamping highlights a key risk for those auditing a blockchain; namely that not all approved transactions may be visible to the auditors. If an

[2] http://bit.ly/1etSTev

authorized party acting maliciously was to generate validly signed transactions from corporate-controlled funds, without broadcasting these to the blockchain, the auditors may be unable to detect their existence if internal processes around signing and auditing access to keys were breached or bypassed. These transactions could then be presented to the network after the fact.

The bitcoin protocol does not feature a per-transaction timestamp, introducing a challenge for auditors attempting to identify all transactions that were generated during a given audit period. There is no time-stamp on transactions, and indeed no way to prevent old transactions from being successfully broadcast on the network and included in a block. Old transactions that fell out of the pool of pending transactions could be later re-broadcast by any party holding a copy of the old transaction, whether maliciously or well-intentioned.

We propose, therefore, that the audit process should also include the movement of all blockchain-based funds between wallets (public keys). This addresses two of the main challenges of the audit: ensuring funds are indeed under control of the organization and preventing historical fraudulent transactions from being re-broadcast in the future. By moving all business funds to a new wallet and address during the process of audit, auditors can be satisfied that the funds are indeed under the control of the organization, since they were transferred to a new account, thus proving the possession of the old private key. By transferring to a new wallet, this transaction will prevent the successful execution of any old, hidden (and thus unaudited) transactions during the previous audit period, since it would be rejected by the network as a double-spend attack, as the funds had already been moved to a new wallet. Secondly, it will ensure that the process of generation of the keys for the new wallet is secure, and compliant with best-practice, for the audit period going ahead, without any transactions generated prior to transfer of funds for future replay.

12. BUSINESS PROCESS

The development of blockchain, distributed ledgers, or indeed any other technology, is done largely to improve the business process. As such, distributed ledgers, at present, are not subject of stand-alone audits. They are, instead, part of a typical corporate audit and thus not done from a technology robustness perspective.

Auditors have an issue with the ephemeral nature of money. Like fiat money, the value of cryptocurrencies relies purely on the value assigned to them by their users. It is not the ability to have a better currency that is the issue, it is the benefit of having it over a distributed computer that is linked into the supply chain. As such, the issue becomes which entity and/or ecosystem is being audited. The audit, in the traditional sense, is no longer appropriate for such an internet-based environment.

When recording the balance of accounts holding crypto-currencies or other such commodities, one accounting challenge faced by auditors is that of ascertaining the currency in which the audit should report the overall balance of funds. While a balance could be reported in the native format of the blockchain-based protocol, this could lead to confusion or uncertainty in future. For example, were blockchain-backed bonds for gold or another physical asset to be used, the audit must highlight that these act as a form of promissory, rather than the tangible asset. In the event of a compromise of the blockchain, or the party holding the assets, the blockchain-backed variant may see a price variation or devaluation due to a lack of confidence, or operation of a fractional reserve process by the physical asset holder.

Where a purely cryptographic currency is involved, the rapid volatility of such cryptocurrencies presents a challenge for audit. While the overall number of coins held may remain constant over a period of time, their value may significantly deviate due to fluctuations in pricing. Due to the relative immaturity of these markets, and the limited liquidity available, there remains the possibility of price and market manipulation. This could potentially be abused by either inside or outside parties for their own financial gain, resulting in a loss to the organization. For example, if an organization placed a stop-loss order on cryptocurrency funds, and a flash-crash was to occur as a result of third-party sell orders lowering the market price of a limited-liquidity commodity, this could lead to a sale being executed, permitting another party to acquire the asset from the stop-loss sale at a preferential price [Chase (2017)]. An audit should, therefore, seek to identify how funds held within exchanges are stored, and whether they are at risk from trading orders such as these, in the event of volatility.

13. THIRD-PARTY HOLDING AND CONTROL

Third-parties always present issues for auditors. Often in a distributed online environment, whether for increased usability or due to shortage of technical skills, funds may be held within potentially insecure wallets, where the private keys are accessible to third-parties. For example, funds may be on deposit with an exchange or other online wallet service. In these circumstances, it may be possible for discrepancies to occur, for example, where the exchange could end up in a deficit as a result of cyber-attack or an insider stealing funds.

Where funds are held by a third party on behalf of the entity being audited, this naturally should raise concerns around the security of those funds; without the private keys being under the control of the organization in question, the funds cannot be accessed in the event of the cessation of service of the third party [Perez (2015)]. This may lead to a material loss and deficit for the organization concerned; consequently, it ought to be recorded during an audit. In addition, where funds are held in a third-party exchange or online wallet, the organization concerned may be unable to demonstrate possession of the cryptographic keys controlling their wallet.

In particular, funds within online exchanges and wallets are often interchanged between accounts without any blockchain-based audit trail. For example, if two users of the same platform transact, this transaction can take place using the exchange software's internal record of balance on each account, avoiding a blockchain transaction being broadcast. In such a scenario, it becomes difficult for an audit to verify the true value of funds within the exchange or wallet, without requiring a full withdrawal to an external wallet where the keys are held by the organization. This would permit identification of the true quantity of funds, and create an auditable blockchain entry showing proof of control of those funds at that point in time.

14. VERIFICATION OF PARTIES

The verification inherent in blockchain presents issues with respect to the audit trail. Blockchain-based transactions occur between public key hashes (addresses) corresponding to cryptographic identities. Best practice in the use of keys dictates that each public key (address) should be used only twice; once to receive funds and once to transfers funds out. The justification for this is that one of the security measures of many blockchain-based currencies, including bitcoin, is designed to conceal and protect the user's public key until a spend transaction is created. Prior to this point, only a one-way derivative of the public key is visible on the blockchain. This means that even compromise of the digital signing algorithms used in bitcoin would not result in a compromise of funds, provided parties follow this guidance.

Where parties do follow this guidance, this creates a challenge for auditors, in that recurring transactions to a recipient will not necessarily (and indeed ideally should not) be directed to the same recipient address. The audit process, therefore, should ensure that the correct recipient was specified, and that the receiving address can be substantiated based upon documentation, such as invoices. Further complicating matters, the private keys used to access a wallet may simply be transferred between parties. This means that an address used to receive legitimate funds by a business could be taken over by a party who was provided these keys by an insider after the funds had been received. This makes it difficult to determine the identity of the party operating an address. The audit process, therefore, should both reconcile recipient addresses against invoices, as well as seek to locate duplicate receiving addresses for scrutiny. In many cases, these may simply be explained by receiving parties using online third-party controlled wallets, or by a party who does not follow the best-practice guidance to use a new receiving address for every transaction. Nonetheless, repeat transactions should be scrutinized to ensure that malicious actors do not attempt to transfer funds to previously-used addresses now under the control of a new beneficiary, for the purpose of money laundering or theft.

15. SMART CONTRACTS AND TIME-LOCKED TRANSACTIONS

Various types of smart contract can exist on blockchains. An auditor needs to look through the code to understand the nature of such contracts [Corin et al. (2005)]. In their simplest form, incoming bitcoin payments can specify cryptographic conditions that must be satisfied before they may be spent, or even processed. For example, a bitcoin transaction may specify a time-lock, such that it will be rejected from the blockchain prior to a certain point in time. Such invalid transactions should not be encountered in the blockchain unless valid, as miners should reject them. Nonetheless, were transactions like this to be discovered due to a software bug in miner validation these blocks would be invalid once the error

was detected, and a chain reversal would occur once miners had been updated to follow the correct rules.

A party being audited may hold non-submitted transactions, signed by parties, promising funds on a time-lock. These should not be considered as valid, however, since the initiating party can reverse these payments by transferring their funds away from the sending address prior to the time-lock condition being satisfied, and the block appearing in the chain. The previously-generated time-locked transaction would now be rejected as invalid due to a double-spend occurring, preventing the recipient from receiving their funds. Consequently, such transactions should be considered, at least from a cryptographic perspective, as little more than a non-binding form of IOU.

16. MULTI-SIGNATURE TRANSACTIONS

Auditors typically check authorized signatories in the physical word. With blockchain, once funds have been received, the unspent transaction output (UTXO), used as the input to a future outbound payment, may specify additional restrictions upon spending. Within bitcoin, these restrictions are relatively constrained, and allow for split-signatures, requiring multiple private keys to be produced in order to spend funds. Funds held under such a system present strong protection against actions by any one individual, although an audit process should still ensure that keys are in place and funds are able to be used (i.e., that keys have not been lost, and funds can still be transferred to a new wallet with split-signature requirements).

An audit should ensure that funds are not held in wallets permitting signatures from any parties that have left the organization, or who should no longer have control of those funds. Even where an N-from-M signature scheme is in use, perhaps requiring two keys from a group of six managers, it is important to audit those who have keys present in the release contract to ensure that two people who have left the organization cannot collude to steal funds prior to a re-keying of the accounts. Since copies may be taken of any keys that are not stored in dedicated hardware security devices, key rotation should take place before a group-based key-holder leaves the organization.

17. ARBITRATION CONTRACTS

The solution we propose for audit is “arbitration-style contracts,” an approach not dissimilar to that proposed by Treleaven and Batrinca (2017). These can be used on UTXOs, to allow two transacting parties to appoint a mutually-agreed arbitrator in a transaction. In such scenarios, the funds may be spent by any two of the three participants (including the arbitrator). Where both transacting parties are in agreement, they may transfer the funds as they wish, since together they hold two of the three keys. Where the two parties enter dispute, the arbitrator can review the circumstances, and sign a judgment that, with the agreement of only one of the parties, will result in the transaction executing. The arbitrator cannot act alone without the consent of one of the parties, since they hold only one of the two required keys to carry out a transaction.

Where such contracts are in use, the audit process should carefully review the contracts in place, and establish the identity of those arbitrating any outstanding transactions. In the scenario where a party to a transaction recommends a non-independent arbitrator, it would be possible for that party to use the corrupt arbitrator to steal funds that the organization under audit may feel they are owed. For this reason, funds that are contract-locked should not be considered to have been received, until a transaction takes place to move them to a wallet under the control of the organization under audit.

18. MICRO-PAYMENT CONTRACTS

Auditors have also got to get used to micro-payments. In some scenarios, where small quantities of funds are being transacted, which would ordinarily be economically infeasible to carry out on the main blockchain, a micro-payment channel can be formed. This is done by the parties in order to permit repeated transactions to take place within the constraints of a larger transaction, which is updated dynamically as transactions take place, altering the funds owed. Under such a scenario, a time-locked transaction is combined with a 2-from-2 multi-signature contract. The end result is that the sending party holds a dual-signed “refund” transaction, granting themselves a full return of the funds paid out, but with a time-lock in place to prevent it from being processed prior to a certain time. A second transaction is then created, forming a “bond” between the two parties. This bond requires both parties to sign to release the funds. Consequently, the initial “refund”

contract can be used (while adhering to the time delay) to return the funds to the initiating party. Only the second "bond" transaction need be produced and transmitted to the blockchain. Similar to the blockchain, as funds are owed to the recipient an updated "refund" transaction is produced and signed by both parties, without the original timelock, allocating the outgoing funds between the two parties agreed. This transaction is again not broadcast to the network, but held by the receiving party. At any point prior to the original time-lock expiring, the receiving party can broadcast their copy of the most recent "refund" transaction to receive the funds they are owed within the micro-payment contract.

Micro-payments are useful for avoiding the large transaction fees on major blockchains, such as bitcoin, and will become an increasing feature of audits going forward. Significant, however, is that the sending party should only engage in such a contract where the transfer of funds is uni-directional; a second contract must be set up if funds may be transferred in the other direction, as otherwise the receiving party could broadcast an outdated version of the release transaction from before a transfer back to the sender. Using two channels, with a clear recipient for each, will avoid this.

Funds within a micro-payment contract should be audited with care, since until the contract completes the exact outcome cannot be certain. If the audited party is the recipient, it is possible no funds will be received if the recipient forgets to broadcast the most recent refund transaction, or broadcasts the wrong refund transaction in error, sending excess funds to the original sender. Likewise, in the event of an outage preventing the recipient from broadcasting their version of the transaction, the sender can broadcast their original refund transaction, and retrieve all of the funds once the time-lock condition is satisfied. Consequently, only micro-payment contracts that have been concluded through the broadcast and inclusion of a release transaction in a block on the chain should be considered to have completed.

We recommend, as a sending party, a micro payment contract should not be considered concluded by an auditor until a refund or release transaction has been made. If no release transaction is made, the refund may be made at any point after the time-lock condition expires; although if this does not occur a release can be made at any point prior to the refund being broadcast, irrespective of time passed.

19. HASHLOCK (PRE-COMMITMENT) CONTRACTS

Hashlocked transactions are another variation of contract-constrained transactions that auditors have to be cognizant of. In hashlocked transactions, a received transaction may only be spent when the corresponding pre-image to a cryptographic hash is provided as part of the transaction. This means that a transaction is created, which specifies that in order to spend the funds produced as an output, it is necessary to provide the input to a one-way function, such that a certain output (contained within the transaction) is yielded. Absent the knowledge of this input value, it is not possible to spend the funds, as the transaction will not be placed into a block by miners. With access to this value, the funds can be spent, as the transaction will be accepted by miners, and included in the blockchain. Once the input value is revealed in a transaction spending the funds held within the hashlock, any party may validate that the transaction is legitimate, by ensuring the hash matches the original requirement.

During an audit, UTXOs protected by a hashlock should be closely reviewed. Without access to the corresponding hashlock release value, funds cannot be spent and are thus inert. Consequently, as with the process detailed earlier for ensuring company funds are genuinely under the control of the entity being audited, an audit should consider whether hashlocked funds are accessible to the organization. Since to demonstrate knowledge of the hashlock key it must inherently be revealed to an auditor, the funds should be transferred to a new hashlock key, thus demonstrating possession and control of the funds, and ensuring that the funds are protected going forwards.

20. CHAIN OBFUSCATION AND COIN MIXING

One potential challenge during an audit is the creation of transactions designed to obfuscate the intentions of the parties making payments, or the handling of coin mixing, in attempts to conceal the trail of transactions. In the first instance, transaction inter-mingling can be used to provide a level of deniability for those making transactions. Using the so-called CoinJoin technique,[3] a contract-based release of coins is used to form a single transaction, incorporating multiple mutually distrusting parties' transactions. Potential participants can create

[2] http://bit.ly/2eRpLU8

a new receiving address for their new coins, and form one transaction between all three parties requiring all participants to sign the transaction to release the funds. Each participant's inputs are then merged in the one transaction, with an output for each party.

The problem with the above weakness for auditors is that it separates the link between the inputs and outputs, since ambiguity is introduced on the blockchain as to which inputs correspond to a given output. If this process were repeated multiple times, blockchain-based analysis to trace funds would be significantly hindered. To establish what happened within each CoinJoin operation, it would be necessary for an auditor to identify and communicate with the other parties in the CoinJoin operation. With no easy way to establish communication with a pseudo-anonymous user of a cryptocurrency, this would be a significant challenge, especially if the process was repeated multiple times.

The technique of mixing or tumbling, while less common due to requiring trust in the provider of the service, is designed to hinder the tracing of transactions involving cryptocurrency coins. A party wishing to "clean" the past history of their coins would transfer these coins to a mixing service as part of a transaction. In return, providing the mixing service is honest, a set of coins would be returned to a new address, which have different origins. Without compromising the mixing service, an audit would be unable to trace funds through a well-implemented mixing service.

During an audit, techniques to obfuscate the true destination or origin of transactions may pose a challenge, as these may hinder the process of confirming that the destination of funds is as stated. For example, an insider attempting to steal company funds would almost certainly attempt to mix their coins using one of these techniques, to avoid their purchases being traceable back to the original theft.

21. CROSS-CHAIN TRANSACTIONS

As a final, almost obvious point, the complexity of the audit increases where more than one cryptocurrency is involved. Different cryptocurrencies may have their own independent blockchains. Where transactions are used to carry out cross-chain trades, these may present a challenge during audit. Such transactions may be encountered when carrying out an exchange between two different cryptocurrencies. For example, if an organization was attempting to trade one cryptocurrency for another, and avoid the risk of the counterparty in the transaction taking their funds without paying the outstanding balance in the other currency, a cross-chain transaction is taking place. Hashlock-type contracts may be used here, since the same hash output value can be used across blockchains, with the corresponding input to unlock the transactions then able to be exchanged, thus making the funds available for release on both blockchains simultaneously. Auditing this transaction would require consideration of both blockchains, potentially significantly increasing the necessary scope of the audit.

22. CONCLUSION

In this paper, we have shown that the audit process, as it currently stands, is not sufficiently robust to handle the challenges of digital money transfer and storage. The questions that auditors need to ask have to change and adapt. Our contribution is in offering some insights into the areas that must be addressed. Specifically, the financial audit must facilitate the distributed nature of blockchain assets, cryptocurrencies, and online legers. The rules and processes that auditors apply need to adapt to the complexity of such distributed systems. In particular, we identify the multijurisdictional nature of digital value and the time stamping of transactions as requiring special attention.

We further illustrated the many weakness and challenges in blockchain, despite the promise of self-audit. These include transaction malleability and both long and short-term blockchain forks. We also demonstrated the challenges presented by DAOs, as well as accounting and auditing cryptocurrencies and distributed ledgers in multiple jurisdictions. We argue that dedicated audit professionals should consider how to address such issues.

In conclusion, we point out that audits are evolving to a more risk-based and distributed model. In this context, distributed ledgers, in effect triple-entry book keeping, present challenges to auditors previously focused solely on double entry book keeping. In this new environment, organizations have multiple counterparties to the same transaction. To address this, we propose that smart contracts be adapted to facilitate self-audit and that the skillset of auditors be adapted to face the new challenges.

References

Allen, R., J. Loebbecke, and K. Sorensen, 1998, "Multilocation audit risks," Journal of Applied Business Research 14:4, 1-13

Andrychowicz, M., S. Dziembowski, D. Malinowski, and Ł. Mazurek, 2015, "On the malleability of bitcoin transactions," International Conference on Financial Cryptography and Data Security, 1-18, Springer

Brink, V., 1988, (Reprint 1941), Modern internal auditing, Ronald Press

Buyya, R., C. Yeo, and S. Venugopal, 2008, "Market-oriented cloud computing: vision, hype, and reality for delivering IT services as computing utilities," in High performance computing and communications, 10th IEEE International Conference on High Performance Computing and Communications (HPCC)

Chase, K., 2017, "GDAX to reimburse traders for losses during flash crash," DEC Brief, June 25, http://bit.ly/2vM97JW

Corin, R., S. Etalle, J. den Hartog, G. Lenzini, and I. Staicu, 2005, "A logic for auditing accountability in decentralized systems," in G. Barthe, A. Datta, and S. Etalle (eds.), Formal aspects in security and trust, Springer

Dusenbury, R., J. Reimers, and S. Wheeler, 2000, "The audit risk model: an empirical test for conditional dependencies among assessed component risks," Auditing: A Journal of Practice & Theory 19:2, 105-117

Francis, J., 2004, "What do we know about audit quality?" The British accounting review 36:4, 345-368

Gervais, A., G. Karame, K. Wüst, V. Glykantzis, H. Ritzdorf, and S. Capkun, 2016, "On the security and performance of proof of work blockchains," Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security

Hegazy, M., and M. Nahass, 2012, "An analysis of multi-location audit risk factors and the improvement of the audit process: an empirical study," Journal of Economics and Engineering 3:1, 35-48

IFAC, 2009, "Handbook of international standards on auditing and quality control," International Federation of Accountants

Nakamoto, S., 2008, Bitcoin: a peer-to-peer electronic cash system, Private distribution

Perez, Y. B., 2015, "Mt Gox: the history of a failed bitcoin exchange," Coindesk, August 4, http://bit.ly/2v0imIO

Rezaee, Z., and A. Reinstein, 1998, "The impact of emerging information technology on auditing," Managerial Auditing Journal 13:8, 465-471

Ringelstein, C., and S. Staab, 2009, "DIALOG: distributed auditing logs," in Web Services, IEEE International Conference

Rogers, V., T. Marsh, and J. Ethridge, 2004, "Internal controls: winning the battle against risks," Internal Auditing-Boston-Warren Gorham and Lamont Inc. 19, 28-34

Treleaven, P., and B. Batrinca, 2017, "Algorithmic regulation: automating financial compliance monitoring and regulation using AI and blockchain," Journal of Financial Transformation 45, 14-21

Whittington, R., and K. Pany, 2012, Principles of auditing and other assurance services, McGraw-Hill Irwin

Targeting the robo-advice customer: The development of a psychographic segmentation model for financial advice robots

DIEDERICK VAN THIEL | AdviceRobo and Tilburg University
W. FRED VAN RAAIJ | Professor of Economic Psychology, Tilburg University

ABSTRACT

The purpose of this study is to develop the world's first psychographic market segmentation model that supports personalization, customer education, customer activation, and customer engagement strategies with financial advice robots. As traditional segmentation models in consumer finance primarily focus on externally observed demographics or economic criteria such as profession, age, income, or wealth, post-hoc psychographic segmentation further supports personalization in the digital advisor's service delivery. It might also provide insight into how to include the 4.5 billion underserved people financially and support inexperienced millennials in securing their future financially. To develop the psychographic segmentation, a survey (N= 2,232) has been conducted across the U.K. and the Netherlands. Factor analysis has been performed to define the following psychographic factors: "convenience," "financial illiteracy," and "rigid personality." Based on these factors, a Ward cluster analysis has been performed to define the psychographic segments across the two markets.

1. INTRODUCTION: THE MARKET OF ROBO-ADVICE

Financial advice is vital for many consumers, as general financial literacy is limited [Van Raaij (2016)]. Financial advice is defined as third-party services that help consumers reach financial decisions [Collins (2010)]. Advice can usually be provided through face-to-face contact, by phone, or digitally. Because today's consumers are asked to make more financial decisions than ever before, as well as the fact that they live in an increasingly complex environment, financial advice is becoming increasingly important [ASIC (2010)]. Financial decision-making, and thus, traditional financial advice, is being transformed by digitalization [Malhotra and Malhotra (2006)]. Younger generations, in particular, specifically millennials, embrace digital lifestyles [PWC (2014)], and such lifestyles have driven the digitization of financial services and the use of financial tools facilitating financial decision-making. As consumers are becoming more and more self-directed, financial capability building is becoming increasingly digital [Van Thiel and Van Raaij (2017)].

Indeed, intelligent agents can be very useful; they are low cost financial assistants [Van Thiel et al. (2008)]. Robo-advisors, intelligent agents, financial assistants, and many other digital decision support systems have the potential to support sound financial decision-making and to reduce financial stress, and thus improve financial security. The ultra-low costs in which these digital advisors can offer their services significantly enlarges the global reach of financial advice. The gap they close is well defined by researchers across the globe. According to the National Financial Capability Study [FINRA (2012)], only 50% of Americans obtained some form of financial advice over a five-year period. According to ASIC, fewer than 40% of the Australian adult population has ever used a financial planner [ASIC (2010)], and in developing countries access to financial advice is even worse. According to the World Bank, 2 billion people lack access to financial services [World Bank (2015)], and many lack access to financial advice to facilitate sound financial decision-making.

Currently, across many geographies, an increasing number of financial service providers are considering the use of robo-advisors: online advice platforms that provide advice by using complex algorithms [Bradbury (2014)]. AT Kearny predicts robo-wealth advisors will manage U.S.$2.2 trillion in assets by 2020 because of the fast-growing adoption rate of this model among younger generations [Van Thiel and Van Raaij (2017)]. The Financial Times estimates that the market for funds advised by hybrid-robo-human services will grow to U.S$16.3 trillion by 2025. Additionally, the growth of the credit-side of the market growth is clear; the global automated peer-2-peer lending market is expected to grow from U.S.$64 billion in 2015 to U.S.$897 billion by 2024 [Transparency Research (2016)].

As the digital lifestyles are becoming commonplace around the world, people are producing an increasing amount of behavioral data. Robo-advisors make use of this and apply algorithms that match consumers or small businesses with financial products or portfolios. As brands increasingly compete on understanding customer DNA at every touch point, behavioral data from all kinds of sources can form the game changers in financial services. Big behavioral data analytics is being applied to find hidden patterns and correlations to more deeply understand customers' behaviors. These deep customer DNA insights can drive personalized and predictive services. Personalization and predictive services are, especially in relation to personal finance and financial planning, services that can facilitate sound financial decision-making across populations.

> "Five million jobs will be lost by 2020 as AI, robotics, nanotechnology and social economic factors replace the need for human workers."
>
> World Economic Forum, Future of Jobs Report, 2016

The increasing number of behavioral data driven services forecast the strong growth of robo-advisors, intelligent agents, and virtual assistants in the coming decades. The University of Oxford places financial advisors on their list of the "Top five jobs that robots are already taking" [Frey and Osborne (2015)]. Frey and Osborne's research indicates that financial analysts and advisors are being replaced by robo-advisors, driven by predictive systems, big data, and computing power. Robo-advice is swiftly growing because of the increasing self-directedness of people; the full potential of robo-advice can only be reached by larger adoption.

The change in customer-need structures through the increasing impact millennials have, as well as the introduction of modern technologies such as digital and

robo-advice, increases the need for more personalized segmentation models. The objective of this research is to develop the world's first psychographic segmentation model that uncovers and monitors the elements that motivate consumers to make use of digital or robo-advice. Psychographic segmentation is an approach to financial advice market segmentation based on the personality characteristics of consumers. Banks can apply it in more personalized strategies for increasing customer engagement and lowering the cost of acquisition and servicing. In this study on financial advice, we focus on one of the most complex forms: mortgage advice. Also, to validate the model, research has been conducted in Europe's two most advanced mortgage advice markets: the U.K. and the Netherlands. The segmentation model provides insight into psychographic factors and variables that drive perceived satisfaction with financial advice robots and can be applied in advanced marketing and risk strategies.

2. FROM CONSUMER SEGMENTATION TO PERSONALIZATION

2.1 Financial advice and customer segmentation models

Product differentiation and market segmentation are marketing strategies aimed at increasing engagement with customers. Product differentiation refers to the differentiation in the product perception due to physical and non-physical attributes, including price, to better meet human expectations [Dickson ad Ginter (1987)]. In imperfect market situations, in occasions where there is no homogeneity for all competitors in the market, market segmentation is another commonly used strategy [Smith (1956)]. Smith illustrated that market segmentation involves viewing a heterogeneous market as several smaller homogeneous markets. Smith suggested three criteria to be fulfilled in segmentation: (1) homogeneity (i.e., communality needs within groups); (2) distinction (i.e., uniqueness between groups); and, (3) reaction (i.e., similarity of response toward marketing strategy, product, offer, or services within a group).

Market segmentation and product differentiation are two sides of the same coin. The link between product differentiation and market segmentation is the product benefit [Van Raaij and Verhallen (1994)]. Market segmentation is also defined as a recognition of the existence of multiple demand functions and the development of a plan to meet one or more of these functions [Frank et al. (1972)]. These groups can be addressed by specially designed, but also standardized, strategies [Kotler & Cox (1980)]. Kotler (1980) claims that market segmentation creates a more finely tuned product or service, offering a price appropriate for the target segment. Kotler (1980) also claims that three major segmentation forms are commonly used: demographic, psychographic, and behavioral [Andreasen et al. (2003)]. Customer segmentation by banks, however, remains largely limited to categories of corporate and retail customers, as traditionally defined [Machauer and Morgner (2001)]. Corporate customers are distinguished by their geographic range of activities, sector, and size. In personal retail banking, externally observed demographic or economic criteria, such as profession, age, income, or wealth are often the preferred dimensions for segmentation [Moutinho and Meidan (1984), Harrison (1994)]. However, demographic and economic criteria are rough indicators for the need structures and the reaction patterns of retail customers [Machauer and Morgner (2001)].

These forms of traditional market segmentation are bound to a high probability that standardized service packages are offered to customers that are not suitable. Thus, low satisfaction and possible migration of customers is to be expected [Machauer and Morgner (2001)]. As segmentation can be forward, backward, and simultaneous [Van Raaij and Verhallen (1994)], modern forms of segmentation are post-hoc. Backward or reverse segmentation means dividing customers into need clusters based on the data a company already has. Thus, a heterogeneous population is surveyed and segments are determined based on homogeneous response patterns from within the population [Machauer and Morgner (2001)]. The research seeks measures that cluster consumers into potentially profitable but unique groups within the population. Some studies in this area use customer responses related to questions on product features or usage. Product usage frequency patterns [Burnett and Chonko (1984)], for example, identified four customer segments for packaging banking products. Accordingly, the segment labels "traditional," "convenience," "investment," and "debt" were derived from the characteristics of the preferred products within these segments.

2.2 Psychographic segmentation

Several other studies using post-hoc segmentation approaches are oriented toward the psychological

determinants of customers, in that they refer to psychographic or benefit segmentation [Machauer and Morgner (2001)]. The purpose of psychographics is to obtain a better understanding of the consumer as a person by measuring psychological dimensions, way of living, interests, and opinions [Ziff (1971)]. The most widely used approach to measure lifestyle is by using activities, interests, and opinion (AIO) rating statements [Plummer (1974)]. A widely used tool for lifestyle segmentation is the VALS scheme [Rokeach (1973)], which blends research of values, hierarchy of needs, and sociology in its operation. Another frequently used tool for lifestyle segmentation is the "list of values" [Kahle et al. (1986)].

In this era, where millennials force their digital lifestyles onto markets and technologies such as machine learning bring enhanced analytic possibilities, big data-based segmentation revolutionizes opportunities for personalized targeting. As, for example, segmentation for mobile devices is typically based on demographics and reported use, smartphone measurement software enables us to directly add observed user behavior and psychographics [Hamka et al. (2014)]. Big data insight is generated through analytics, which can be subdivided into descriptive analytics (analytics activities that explain the past), predictive analytics that predict/forecast future outcomes, and prescriptive analytics, which predict future outcomes and suggest options for decision-making. In the last step of the "virtual value chain," the data or insight might be visually represented, the data distributed, or access to the data or analytics results provided, for example, through an API [Hartmann et al. (2014)].

Hypothesis 1: For defining digital or robo financial advice market segments, psychographic segmentation can be applied for developing personalized robo-advice strategies.

2.3 Financial advice and financial literacy

On the demand side of robo and digital financial advice, the issue in the global financial advice markets is that financial advice is being perceived as inaccessible. The National Financial Capability Study [FINRA (2012)] conducted a survey among U.S. citizens (n=1488) to discover the number who obtained some form of financial advice in the past five years and found that 57% had. 8% had received advice on debt management, 21% on tax planning, 24% on receiving a loan, 33% on investments, and 33% on insurance.

Furthermore, the survey showed that a higher income, good education, and sound financial literacy led to more financial advice being sought. Also, the generation of millennials are approaching their peak buying years [Goldman Sachs (2017)]. According to Goldman Sachs this generation is on their way to marry, buy cars and houses for their family life. PWC studies the financial literacy of millennials [PWC (2015)] and found that millennials struggle with their financial lives. Only 24% have adequate financial knowledge, 34% are very dissatisfied with their current financial situation, but still only 27% seek professional financial advice on saving and investment. They are inexperienced financially and also embrace a digital lifestyle, thus new financial technology such as financial advice robots can fill this knowledge gap.

Hypothesis 2: Financial illiteracy is a differentiating factor in the psychographic segmentation model for developing personalized robo-advice strategies.

2.4 Financial advice and motivation

FINRA (2012) found that people's level of financial literacy and education impacts their openness to financial advice. PWC research illustrates that only 27% of millennials use financial advice. Motivation to use digital financial advice, therefore, is a potential factor that differentiates customer segments for building personalized robo-advice strategies. To be motivated means to be moved to do something;

motivation concerns energy, direction, persistence, and equifinality; all aspects of activation and intention [Deci and Ryan (2010)]. Consequently, a person who feels no impetus or inspiration to act is characterized as unmotivated, whereas someone who is energized or activated toward a goal is motivated. People do not only have different degrees of motivation, but also various kinds of motivations. The orientation of motivation concerns the underlying attitudes and goals that give rise to action [Deci and Ryan (2010)]. In their "self-determination theory," Ryan and Deci distinguish several types of motivation based on various reasons or goals that give rise to an action. Whereas intrinsic motivation refers to doing something because it is inherently interesting or enjoyable, extrinsic motivation refers to doing something as a means to reach another goal or outcome.

Hypothesis 3: Motivation is a differentiating factor in the psychographic segmentation model to develop personalized robo-advice strategies.

2.5 Financial advice and risk appetite

Nelson (1970) classifies products into search and experience goods. Search goods offer consumers the ability to obtain product quality information prior to purchase, whereas experience goods, like financial advice, do not. Credence goods are a specific category of experience goods. Wolinsky (1995) defines credence goods as experience goods whose sellers are also experts who determine customers' needs. Information asymmetry in credence goods markets lead to prices that embody mark-ups over costs. Furthermore, the equilibrium does not maximize expected customers' surplus. Another consequence of information asymmetry in credence goods is fraud [Wolinsky (1995), Emons (1997)]. Since customers can never be certain about the quality of the sellers' services, experts have opportunities and incentives to cheat. Consequently, regulators force financial advisors to investigate the risk profile of a customer and match their advice. Due to the 2008 economic crisis, trustworthy assessments of risk perception and risk tolerance of financial customers became a central element in financial supervision [Roszkowsky and Davey (2010)]. To differentiate between risk tolerance and risk perception, we must first define risk. According to Roszkowsky and Davey, risk is the uncertainty that exists as to what the eventual outcome will be. Risk arises in any decision where there is some doubt about at least one of the possible outcomes. The risk inherent in any given situation will depend on the range of possible outcomes and the likelihood and value of each outcome. Thus, in a financial context, risk tolerance is the amount of risk an individual chooses when making a financial decision. Although risk tolerance is largely a fixed personality trait and stable, it is nonetheless marginally subject to situational influences (for example mood) and may change due to life circumstances (for example aging).

Furthermore, an evaluation of the degree of risk generally involves a perception of the situation, which means that there is some interpretation of the objective reality. Tversky and Kahneman (1974) show that risk perception is a function of intuitive notions of risk (e.g., probability of loss and loss aversion) rather than of technical risk measures such as beta, standard deviation, or variance.

Hypothesis 4: Risk tolerance is a differentiating factor in the psychographic segmentation model to develop personalized robo-advice strategies.

2.6 Purpose of this research

As defined in the introduction, the purpose of this study is to develop the world's first psychographic market segmentation model that supports personalization, customer education, customer activation, and customer engagement strategies with financial advice robots. As traditional segmentation models in consumer finance primarily focus on externally observed demographics or economic criteria such as profession, age, income, or wealth [Meidan (1984), Harrison (1994)], post-hoc psychographic segmentation might support further personalization in financial advice robot service delivery. It might furthermore provide insight on how to include the 4.5 billion underserved people financially and support inexperienced millennials with building financially stable lives.

3. METHODOLOGY AND RESULTS

3.1 Developing the "digital psychographic segmentation" (DPS) model

3.1.1 DEVELOPING A CONCEPTUAL MODEL AND SURVEY DESIGN

Firstly, the main psychographic dimensions were determined. These psychographic dimensions served as the conceptual basis for a questionnaire (which is explained in the following section) that aimed to gain

evaluations from digital mortgage advice customers of their psychographic profile and their perceived acceptation of digital financial advice. Based on earlier scholarly research [Machauer and Morgner (2001), FINRA (2009), Deci and Ryan (2010), Tversky and Kahneman (1975)], AFM field studies [Van Raaij (2016)], and several brainstorming sessions among academic peers, the variables for psychographic segmentation have been defined.

The survey confirmed the importance of the following variables:

1. Need for relevant information
2. Time spent on finding relevant information
3. Span of alternatives evaluated in the decision process
4. Support being asked for in the information search
5. Level of trust in financial advisors
6. Openness to new products and services
7. Rationality in the buying decision
8. Level of financial knowledge
9. Level of following financial market developments
10. Risk appetite
11. Level of maximization in the buying decision
12. Level of "social opinions" applied in buying decision

The psychological profiles composed of these variables should characterize the way a person is open to digital financial advice. It should provide an input to digital and robo-advice leadership teams on how to design digital advice strategies and personalized services per customer segment.

Based on the variables, a survey was composed in close cooperation with research agency, GfK. The survey design consisted of five sections with 38 questions on satisfaction with attributes of the advisor and the advice service quality. Section A contained screening questions to build a social-demographic profile of the respondents. Section B contained questions on the performance evaluation of recent mortgage closings. Respondents were asked to rank their experience with the performance on the variable list of their most important financial advisor. To measure latent constructs such as opinions and interests, scoring took place on a 7-point Likert-scale. Section C contained questions for benefit evaluation on the variable list of mortgage closings. Again, scoring took place on a 7-point Likert-scale. These sections were applied in composing the DCX-model [Van Thiel and Van Raaij (2017)]. Sections D and E contained statements about respondents' psychographics. To measure the latent construct of financial behavior, the responses were also measured on a 7-point Likert scale. For this paper on psychographic segmentation, Section E and the response in Section A are used. The survey was validated on a pilot group of 100 respondents prior to the larger online field experiment.

"85% of millennials prefer using robo-advisors over traditional advisors"

Charles Schwab, 2017

3.1.2 PERFORMING THE SURVEY

To develop and test the model, an online survey of 2,332 consumers was conducted across two experiments. Respondents were grouped based on their recent experience with buying a house and choosing a mortgage. The respondents were randomly selected from the GfK-consumer panels of the U.K. and the Netherlands and divided into two groups (1) 2013 the Netherlands (n=1407); and (2) 2013 The U.K. (n=935). The first online survey experiment was held in Q1-2013 in the Netherlands (n=1407). In the experiment, experienced financial advice users (N=815) were differentiated from the inexperienced users (n=592). The second experiment was held in Q4-2013 in the U.K. (n=935). Because the differentiation between experienced and inexperienced users in the first experiment created no significant insights, the respondents in this experiment were only experienced advice users (consumers who had received mortgage advice in the last year). To be able to compare the results over the experiments, the same questionnaires and analyses were applied. The experiments were used to develop and validate the DPS model.

3.1.3 PSYCHOGRAPHIC SEGMENTATION

On the survey response, principal component factor analysis with Varimax rotation was applied with the 12 psychographic variables. The Varimax rotation with Kaiser normalization was performed over four rotations. Both in the U.K. as in the Netherlands, three components (convenience with eigenvalue of 1.89; financial illiteracy with eigenvalue of 1, and rigidity with eigenvalue of

0.51) resulted from the principal component analysis. Based on the Kaiser criterium that eigenvalues >1 should be selected, only convenience and financial illiteracy should be components in the psychographic model. However, based on the importance of risk appetite and emotionality in decision-making according to regulators, a third component "rigidity" is added to the model. The rotated component matrix for the U.K. is shown in Table 1 and for the Netherlands in Table 2.

With the components discovered, a Ward's minimum variance clustering model was applied to classify financial advice consumers. Based on the three components, four dominant segments were obtained explaining psychographic differences in suitability for digital financial advice. The following decision rules were applied to assign to each cluster:

1. Explained variance
2. Group size >100
3. Interpretation of homogeneity of the groups
4. If no clear interpretation, start at rule 1

Table 1: Rotated component analysis U.K.

	ROTATED COMPONENT MATRIX[a]			
	COMPONENT			
	CONVENIENCE		FINANCIAL ILLITERACY	RIGID
Information: Seek a lot – Try to limit	e18_1	.800		
Time: Take all I need – As quickly as possible	e18_2	.737		
Alternatives: Many – Limited amount	e18_3	.801		
Research: All myself – Let others do as much as possible	e18_4	.646		
Trust advisors: Easily – Do not	e18_5			.671
Products: Try new – Stick to known	e18_6	.407		.468
Decisions: Based on feelings – Logically and systematically	e18_7			.747
Financial knowledge: A great deal – Very little	e18_8		.832	
Financial market developments: Fully abreast – Barely follow	e18_9		.834	
Risks: Fully prepared to take – Averse (maximum security)	e18_10		.598	.453
Product search: Until the best possible – As soon as found	e18_11	.826		
Comparitive shopping (usage rating & reviews websites): Use a lot – Do not use	e18_12	.628		

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in four iterations.

Table 2: Rotated component analysis the Netherlands

	ROTATED COMPONENT MATRIX[a]			
	COMPONENT			
	CONVENIENCE		FINANCIAL ILLITERACY	RIGID
Information: Seek a lot – Try to limit	e18_1	.830		
Time: Take all I need – As quickly as possible	e18_2	.776		
Alternatives: Many – Limited amount	e18_3	.701		
Research: All myself – Let others do as much as possible	e18_4	.642	.420	
Trust advisors: Easily – Do not	e18_5		-.464	.394
Products: Try new – Stick to known	e18_6			.719
Decisions: Based on feelings – Logically and systematically	e18_7			.645
Financial knowledge: A great deal – Very little	e18_8		.780	
Financial market developments: Fully abreast – Barely follow	e18_9		.726	
Risks: Fully prepared to take – Averse (maximum security)	e18_10		.401	.677
Product search: Until the best possible – As soon as found	e18_11	.808		
Comparitive shopping (usage rating & reviews websites): Use a lot – Do not use	e18_12	.628		.357

Extraction method: principal component analysis. Rotation method: varimax with Kaiser normalization.
a Rotation converged in four iterations.

Table 3: U.K. psychographic segments and means

		F1: CONVENIENCE	F2 FINANCIAL ILLITERACY	F3 RIGID
		MEAN		
WARD METHOD	C1 Convenience	1.56149	-.05108	.09060
	C2 Trustful with knowledge	-.26301	-.90491	-.84376
	C3 Rigid	-.44974	-.33342	1.03366
	C4 No financial knowledge	-.28716	.83885	-.40067

Table 4: The Netherlands psychographic segments and means

	CONVENIENCE	TRUSTFUL WITH KNOWLEDGE	RIGID	NO FINANCIAL KNOWLEDGE
	1	2	3	4
	MEAN	MEAN	MEAN	MEAN
Convenience	1.6	-0.3	-0.4	-0.3
Financial illiteracy	-0.1	-0.9	-0.3	0.8
Rigid	0.1	-0.8	1.0	-0.4

The dominant psychographic segments and their means are displayed in Tables 3 and 4.

The validation approach for the discovered psychographic segments was in performing this study across Europe's two foremost mortgage advice markets: The U.K. and the Netherlands. The U.K. and the Netherlands are understood as front-runners because of their advanced financial advice eco-systems (financial advisors, governmental regulation, and fintech industry).

3.2 The DPS-model and its validation in two markets

3.2.1 DIGITAL PSYCHOGRAPHIC SEGMENTATION IN THE NETHERLANDS

The principal component analysis in the first experiment shows the factors and their variables across the experienced and inexperienced customer groups through the surveys in the U.K. and the Netherlands that point toward similar factors for psychographic segmentation in digital financial advice. This supports the explanation of the cross-cultural expectations of digital financial advice. The factors are (1) need for convenience, (2) level of financial literacy, and (3) need for rigidity. The DPS-model is presented in Figure 1.

Although the inexperienced and experienced people show great similarities, the factor impact differs across both customer groups. Inexperienced people give more importance to trust in advisors (24.7%) and process rigidity (18.4%), whereas experienced people give more importance to financial knowledge (34.2%).

Applying the factors in a Ward cluster analysis yielded a 4-segment psychographic model. For defining the psychographic profiles, the answers in Section E of the questionnaire were merged with Sections A, B, and C. The profiles designed are: (1) convenience seekers; (2) trustful with knowledge, (3) rigid, and, (4) financially ignorant (no financial knowledge). The psychographic segments are shown in Figure 2.

Convenience seekers score high on (1) trying to limit information search (.830), (2) buying the product as soon as one is found (.808), (3) buying the mortgage as quickly as possible (0.776), (4) evaluating limited alternatives (.701), and (5) letting others do the research (.642).

Rigid people are more conservative mortgage buyers. They do not trust advisors very much (.394), tend to stick to known products and brands (.719), decide logically and systemically (.645), and are risk-averse (.677).

On the knowledge factor, differentiation is seen in the segments "trustful with knowledge" with high scores on financial knowledge (.780) and financial market developments (.726). The knowledge aspect makes this segment progressive, showing a willingness to spend time on finding relevant information and the best product for their needs. The customer segment "financially ignorant" instead spends no time on financial decision-making. They, therefore, tend to be neutral and follow financial advice.

Figure 1: Digital psychographic segmentation model

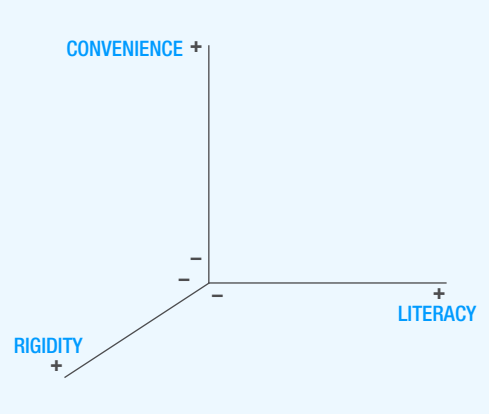
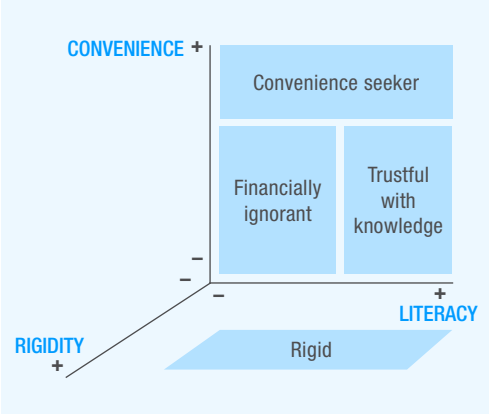


Figure 2: Digital psychographic consumer segments



3.2.2 DPS IN THE U.K.

Although the second experiment (U.K.) also shows the same psychographic factors segmenting the market, their impact differs from the Dutch market. Financial knowledge (34.1%) is the most important variable. But rigidity (28.1%) is important for the British. Convenience (17.1%), on the other hand, is less important in the U.K. Applying the factors in a Ward cluster analysis led to a three-segment psychographic model. The British psychographic customer segments resulting from the Ward-analyses are: (1) convenience seekers, (2) the financially illiterate, and (3) rigid consumers.

As in the Netherlands, British convenience seekers score high on (1) trying to limit information search (.800), (2) buying the product as soon a one is found (.826), (3) buying the mortgage as quickly as possible (0.737), (4) evaluating limited alternatives (.801), and (5) let others do the research (.646). Different from the Dutch, British convenience seekers do look at comparative shopping instruments such as ratings and reviews (.628).

Rigid people are more conservative mortgage buyers. They do not trust advisors very much (.671) and decide logically and systemically (.747) They are less loyal than the Dutch mortgage buyers in their intention to remain loyal to known products and brands (.468).

The financially illiterate group are the third psychographic segment in the British market with low financial knowledge (.832), low financial market knowledge (.834), and are looking for financial security (.598).

3.2.3 HYPOTHESIS TESTING FOR MODEL INTEGRATION

Hypothesis 1: For defining digital or robo financial advice market segments, psychographic segmentation can be applied for developing personalized robo-advice strategies.

The principal component analysis with Varimax rotation showed three factors that differentiate in psychographic mortgage advice segments. The factors are convenience (eigenvalue of 1.89), financial illiteracy (eigenvalue of 1), and rigidity (eigenvalue of 0.51). Based on the Kaiser criterium that eigenvalues >1 should be selected, only convenience and financial illiteracy should be components of the psychographic model. Consequently, Hypothesis 1 can be validated.

Hypothesis 2: Financial illiteracy is a differentiating factor in the psychographic segmentation model for developing personalized robo-advice strategies.

The principal component analysis with Varimax rotation showed that financial illiteracy is a differentiating factor in psychographic advice segmentation. The eigenvalue of 1 is good enough to select financial illiteracy as a factor. Financial illiterates in the U.K. have low financial knowledge (.832), low financial market knowledge (.834), and are looking for financial security (.598). In the Dutch market, financial illiterates have low financial knowledge (.780) and financial market knowledge (.726). Also, the study in the Dutch market revealed a fourth segment “trustful with knowledge.” They are the most progressive customer segment and spend a lot of time finding the best solution matching their needs. Financial (il)literacy is a differentiating factor in psychographic customer segmentation. Hence, the second hypothesis can also be validated.

Hypothesis 3: Motivation is a differentiating factor in the psychographic segmentation model to develop personalized robo-advice strategies.

The principal component analysis did not show motivation as a differentiating psychographic factor. Nevertheless, in the factor, scores differentiating between people who actively seek relevant information, spend time and keep searching until they find the best solution for their needs show high loadings. These people are segmented in the U.K. in the psychographic segment of the "rigids." In the Netherlands, they are segmented in the psychographic segment of "trustful with knowledge." Hence, although not a factor in the model, someone's personal motivation is something to consider when developing digital strategies. Thus, Hypothesis 3 cannot be validated.

Hypothesis 4: Risk tolerance is a differentiating factor in the psychographic segmentation model to build personalized robo-advice strategies.

The principal component analysis did not show risk tolerance as a differentiating psychographic factor. The influence of risk tolerance in the U.K. is seen in the customer segment of the financially illiterate who seek maximum security. In the Dutch psychographic segmentation risk tolerance also influences the behavior of the financially ignorant who seek maximum security. But also, "trustful with knowledge" with a higher risk tolerance include people that are open for trying new products and services. Nevertheless, Hypothesis 4 cannot be validated.

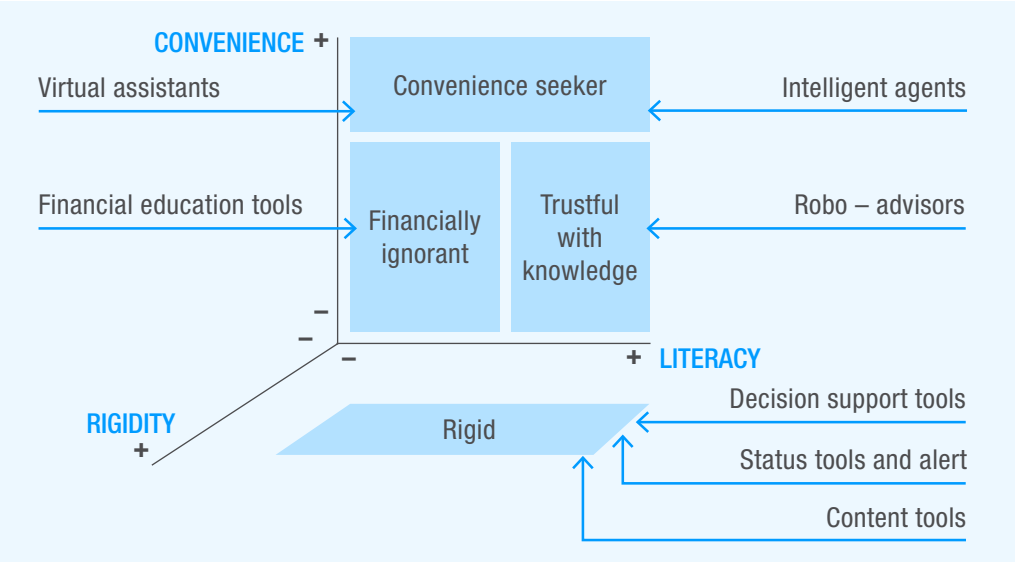
4. DISCUSSION

As defined in the Introduction, the purpose of this study is to develop a world first psychographic market segmentation model that supports personalization, customer education, customer activation, and customer engagement strategies with financial advice robots. As traditional segmentation models in consumer finance primarily focus on externally observed demographics or economic criteria, such as profession, age, income, or wealth [Meidan (1984), Harrison (1994)], this research discovered that post-hoc psychographic segmentation supports further personalization in banking advice services.

The cross-cultural study discovered three differentiating psychographic factors that should always be considered when targeting a market for digital or robo-advice. Convenience appears to be the dominant factor and goes back to the need to buy the best fit product with the lowest effort. Content personalization strategies, as well as predictive, prescriptive, and automated services will appeal to this psychographic segment. When executed well, this segment is approachable for financial intelligent agents and virtual assistants.

Financial illiteracy is the second psychographic factor to consider when targeting a digital advice market. Financial illiteracy is bipolar. Those with low literacy need education and support in their financial decision-making. Their overarching need is to find solutions that provide financial security. The financially ignorant

Figure 3: Digital Psychographic Consumer Targeting



are not the most promising for digital or robo-advice. Nevertheless, digital advisors can support customers in this segment with education, orientation, and transaction tools that underpin their financial decision-making. The financial inclusion of the 4.5 billion globally underserved people is an opportunity to develop these kind of digital education tools.

The other pole is the customers with high financial literacy. Customers in this segment have a high interest in personal finance and market developments. They also tend to spend a lot of time in developing their knowledge and are open to new products and services. Because of their knowledge, they are often role models for their friends and families. This highly literate customer segment is appealing for financial advice robots to approach. In parallel with the convenience segment, content personalization strategies and predictive services will appeal to this segment. Different from the convenience segment, the "trustful with knowledge" customers appreciate additional information like reviews, ratings, and blogs to grow their knowledge.

The third psychographic factor to consider when targeting a market for digital advice is the rigidity of people. Rigid people are conservative in their decision-making and want to be in control. They tend to spend a lot of time on finding the best solution because they do not really trust advisors. This segment seems unappealing for financial advice robots because of the trust element. However, if digital tools like, for example, virtual financial assistants are developed that smartly support information searches and improve financial control, they might want to use them. For digital and robo-advisors a targeting model is presented in Figure 3.

Banks and regulators can apply the DPS-model in their service development or supervision. Banks can apply the DPS-model in personalized tools and treatments to their vulnerable customers. Active servicing strategies to avoid bad payment behavior, but also personalized collection strategies, can be encouraged by the regulators applying the DPS-model. In addition, through smart targeting of customer segments with financial advice robots, DPS provides valuable insights. Convenience seekers and people with high financial knowledge are the preferred target groups. For the financially ignorant with rigid behavior patterns, financial robots should focus on education, insight, and control. Furthermore, based on psychographic segmentation, regulators can validate to what extend banks support a customer's personal decision-making. Are the illiterate customers getting the right educational tools and treatment? Are the rigid customers exposed to the right information and control tools, and are robo-advice solutions providing enough information to support sound financial decision making?

The model can also inspire researchers studying changing consumer needs to improve their contribution to closing the advice gap. Applied research on big data-driven robo-advice, virtual assistants, alternative risk modeling, and emerging forms of data-driven financial education and advice can significantly contribute towards closing the financial advice gap. As of 2012, about 2.5 exabytes of data are created each day, and that number is doubling every 40 months – 90% of global data today was produced in the past two years. More data cross the internet every second than were stored in the entire internet just 20 years ago. Big data is arriving from multiple sources with an alarming velocity, volume, and variety. This data will open new opportunities for financial advice robots to personalize their content and services on the psychographics and behaviors of their customers. New artificial intelligent techniques like machine learning make it possible for digital advisors to apply these unstructured data to personalize customer experiences.

5. LIMITATIONS

There are some limitations to this research that should be considered when interpreting the findings. The new models are derived from research in the U.K. and Dutch markets. Although these markets are known as precursors in digital financial advice, additional geographic-specific research must be conducted to generalize the findings to other global markets.

The data of this research were gathered in the year when both the U.K. and the Dutch governments introduced commission stops on complex financial advice. This has stimulated the growth of online low-cost financial advisors since 2013. Due to the brief period between the introduction of the commission stop and when this research was conducted, the price and accessibility benefits offered by these online financial advisors are still in their infancy. Additional time-series research should be performed to monitor the changing impact of the drivers in the DPS-model.

Finally, this research has been singularly performed on digital mortgage advice. Further research can be conducted to focus on business models such as comparison sites, robo-wealth advisors, and advisors in other service industries such as healthcare and utilities.

References

Andreasen, A. R., P. Kotler, and D. Parker, 2003, Strategic marketing for nonprofit organizations, Prentice Hall

ASIC, 2010, "Access to financial advice in Australia," Australian Securities & Investment Committee, Report 224

Bradbury, T., 2014, "Robo advice is coming: what it means, who will buy it-and why," Professional Planner 69, 40

Burnett, J. J. and L. B. Chonko, 1984, "A segmental approach to packaging bank products," Journal of Retail Banking 6.1/2, 9-17

Collins, M. J., 2010, "A review of financial advice models and the take-up of financial advice," Center for Financial Security WP 10.5

Demirguk-Kunt, A., L. Klapper, D. Singer, and P. Van Oudheusden, 2014, "Measuring financial inclusion around the world," World Bank Policy Working Paper

Dickson, P. R., and J. L. Ginter, 1987, "Market segmentation, product differentiation, and marketing strategy," Journal of Marketing 51:2, 1-10

Emons, W., 1997, "Credence goods and fraudulent experts," RAND Journal of Economics 28:1, 107-119

FINRA, 2012, "Financial capability in the United States," Report of findings of the 2012 national financial capability study

Frank, R. E., W. F. Massy, and Y. Wind, 1972, Market segmentation, Prentice-Hall

Frey, C. B., and M. Osborne, 2015, "Technology at work: the future of innovation and employment," Citi GPS

Goldman Sachs, 2017, "Millennials coming of age," Goldman Sachs

Hamka, F., H. Bouwman, M. Reuver, and M. Kroesen, 2014, "Mobile customer segmentation based on smartphone measurement" Telematics and Informatics 31:2, 220-227

Harrison, T. S., 1994, "Mapping customer segments for personal financial services," International Journal of Bank Marketing 12.8, 17-25

Hartmann, P. M., M. Zaki, N. Feldmann, and A. Neely, 2014, "Big data for big business? A taxonomy of data-driven business models used by start-up firms," A taxonomy of data-driven business models used by start-up firms

Kahle, L. R., S. E. Beatty, and P. Homer, 1986, "Alternative measurement approaches to consumer values: the list of values (LOV) and values and life style (VALS)," Journal of consumer research 13:3, 405-409

Kotler, P., 1980, Principles of Marketing, Prentice-Hall

Machauer, A. and S. Morgner, 2001, "Segmentation of bank customers by expected benefits and attitudes," International Journal of Bank Marketing 19:1, 6-18

Malhotra, R., and M. K. Malhotra, 2006, "The impact of internet and e-commerce on the evolving business models in the financial services industry," International Journal of Electronic Business 4:1, 56-82

Meidan, A., 1984, Macmillan studies in marketing management, Macmillan

Moutinho, L., and A. Meidan, 1989, "Bank customers perceptions, innovations and new technology," International Journal of Bank Marketing 7:2, 22-27

Niemiec, C. P., R. M. Ryan, and E. L. Deci, 2010, "Self-determination theory and the relation of autonomy to self-regulatory processes and personality development," in Hoyle, R. H. (ed.), Handbook of personality and self-regulation, Blackwell Publishing.

Nelson, P., 1970, "Information and consumer behavior," Journal of political economy 78:2, 311-329

Plummer, J. T., 1974, "The concept and application of life style segmentation," Journal of Marketing 38:1, 33-37

PWC, 2014, "Engaging and empowering millennials," A follow-up to PWC's next-gen global generational study, PriceWaterhouseCoopers

Rokeach, M., 1973, The nature of human values, Free press

Roszkowski, M. J., and G. Davey, 2010, "Risk perception and risk tolerance changes attributable to the 2008 economic crisis: A subtle but critical difference," Journal of Financial Service Professionals 64:4, 42-53

Smith, W. R., 1956, "Product differentiation and market segmentation as alternative marketing strategies," Journal of Marketing 21:1, 3-8

Transparency Research, 2016, "Increasing small business units to act as building blocks for peer to peer lending market," Transparency Research

Tversky, A., and D. Kahneman, 1974, "Judgment under uncertainty: heuristics and biases," Science 185:4157, 1124-1131

Van Raaij, W. F., 2016, Understanding consumer financial behavior. Money management in an age of financial illiteracy, Palgrave Macmillan

van Raaij, W. F. and T. M. M. Verhallen, 1994, "Domain-specific market segmentation," European Journal of Marketing 28:10, 49-66

van Thiel, D., J. Goedee, and W. J. M. Reijnders, 2008, "Riding the waves. The future of banking starts now," Pearson Prentice Hall

van Thiel, D., and W. F. Van Raaij, 2017, "Explaining customer experience of digital financial advice," Economics 5:1, 69-84

Wolinsky, A., 1995, "Competition in markets for credence goods," Journal of Institutional and Theoretical Economics (Zeitschrift für die gesamte Staatswissenschaft) 151, 117-131

World Bank, 2015, "Access to financial services – a key to lifting people out of poverty.' World Bank research

Ziff, R., 1971, "Psychographics for market segmentation," Journal of Advertising Research 11:2, 3-9

BUSINESS MODELS

PREVIOUS EDITIONS OF THE CAPCO JOURNAL OF FINANCIAL TRANSFORMATION ARE AVAILABLE AT **WWW.CAPCO.COM/INSTITUTE**

Avoiding pitfalls and unlocking real business value with RPA

The impact of financial regulation on business models of cooperative banks in Germany

Transforming the theory and practice of risk management in financial enterprises

Reconciliations: Five trends shaping the future landscape

Thank you and goodbye – ending customer relationships and its significance

Avoiding pitfalls and unlocking real business value with RPA

LAMBERT RUTAGANDA | Consultant, Capco
RUDOLF BERGSTROM | Senior Consultant, Capco
AVIJEET JAYASHEKHAR | Managing Principal, Capco
DANUSHKA JAYASINGHE | Associate, Capco
JIBRAN AHMED | Managing Principal, Capco

ABSTRACT

Financial institutions have quickly adopted robotic process automation (RPA) in recent years, owing to the plethora of manual operational processes in the industry, in order to benefit from significant opportunities for cost reduction and efficiency gains. It is quite unlikely that any executive would say no to a relatively cheap solution that enables the workforce to focus on more value adding activities and adjusts the overall cost structure. Whilst the RPA market growth rate paints a bright future, early RPA adopters have reported challenges with meeting expectations at the outset. Many tall claims have been made of reducing costs, increasing accuracy, improving compliance, and automating work at a fraction of time and cost compared to typical IT projects. However, the high expectations of RPA have created confusion amongst business stakeholders with regards to the capabilities, benefits, and use cases of RPA tools.

Understanding RPA suitability and quantifying the associated benefits is challenging, as many organizations view it primarily as a cost reduction tool, limiting the scope and benefits that can be realized. This has further contributed to the unrealistic expectations business stakeholders have outlined for RPA to deliver and resulted in over a third of RPA projects failing to deliver those expected benefits. Thus, when embarking on an RPA journey one needs to first determine which use cases are suitable for RPA within the context of the associated benefits, as well as the potential drawbacks or pitfalls. In this paper, we explore how RPA has been used across various industries, the challenges faced by early adopters, and the approaches to overcoming these challenges to ensure the real potential of RPA can be unlocked by financial services organizations.

1. INTRODUCTION

Since the early 1990s, financial institutions have been increasing investments in technology and process improvements to harness economies of scale and drive cost efficiencies. Despite these investments, financial institutions continue to struggle to respond to changing customer needs, running efficient and cost-effective operations on legacy infrastructures, and adhering to the wave of new regulatory requirements and security standards. Existing business and operating models are also being disrupted through exponential growth in computational power, technology advancements, and new market entrants, thus increasing the challenge of remaining relevant and competitive.

The ever-evolving marketplace in which financial services firms operate has resulted in a strategic quest for automating and streamlining functions and processes, with improved cost structure, efficiency, and quality as ultimate objectives. At the beginning of 2000, different methodologies, such as lean six sigma, process optimization, and change management tools, as well as offshoring and nearshoring initiatives, were pursued to reduce costs and streamline operations. These pursuits fueled the automation agenda for many organizations, albeit not fully eradicating manual and repetitive tasks performed by humans.

In recent years, RPA has entered the public domain with the promise of helping firms in the quest for automation. RPA has attracted a lot of attention, and even adoption, across the financial services industry, as hopes are that manual and repetitive tasks performed by humans can be automated to improve the overall service delivery, whilst at the same time reducing operational costs, with some studies putting that figure in the 40% to 75% range [Infosys Consulting (2017)].

One of the major benefits of RPA is that it is able to automate business processes within existing applications and technology infrastructures, hence limiting the impact on existing IT architecture. Furthermore, technological innovations, such as RPA, also encompass a number of steps required to enhance cognitive solutions, such as "artificial intelligence" (AI), and machine learning, which, once matured, will further transform the automation agenda in the financial services industry.

According to McKinsey (2017), 16% of available working hours across industries in the U.S. are spent on data processing and 17% on data collection, which are activities that are mainly performed by humans. The automation potential within the respective activities measured as percentage of time spent, is 69% in data processing and 64% in data collection, which when combined means that one-third of the available working hours in the future have more than 60% automation potential.[1]

What is RPA?

RPA is the use of software or "robots" to mimic actions a human user would perform on a computer, at scale, in order to automate the human element of the mundane, manual, and repetitive tasks. By definition, it allows humans to become more human at work. RPA tools integrate with existing applications to interpret interfaces, manipulate data, trigger responses, and communicate across multiple systems. RPA tools seek to automate business processes that are highly repetitive, rule-based, and use structured data to make them more repeatable, faster, and less prone to human errors. The key differentiators between other automation options and RPA is the approach of emulating human actions through a standard user interface, coupled with simple integration with existing applications, requiring limited to no modifications.

There are various RPA tools, ranging from solutions that can handle single transactions from an individual desktop with limited capability of handling different data sources, to those that can manage multiple business processes simultaneously from enterprise servers. Opportunities for RPA in major organizations are many and vary depending on the circumstances. When integration across systems or automation alternatives are too expensive or time-consuming, RPA may serve as a good interim solution. RPA can also be considered in-lieu of outsourcing, as well as other process change management and optimization initiatives.

Vishnu et al. (2017) present a number of frameworks for identifying and evaluating candidate RPA functions. In order to determine the best uses of RPA, the authors also presented a conceptual framework based on the velocity of business change and the stability of IT systems, which determines the applicability of RPA. In essence, RPA works best when the velocity of business change is low and changes to underlying systems are infrequent. For example, a very simplistic and suitable

[1] Chetan Dube, Founder, IPsoft Inc., recently stated that: "The last decade was about replacing labor with cheaper labor. The coming decade will be about replacing cheaper labor with autonomics."

candidate for RPA is static data processing within systems that do not change frequently.

The highest adoption rate of RPA tools has been within the business process outsourcing (BPO) industry, as many of these processes are performed within rigid legacy systems that are not updated on a regular basis and the input data is relatively static.

Nevertheless, the potential benefits of implementing RPA are many, including cost reductions, improved efficiency and quality, faster output, and the ability to integrate with legacy systems. Vishnu et al. (2017) also outlined examples of core RPA benefits ranging from improved operational agility, scalability and compliance, business planning and forecasting, to enhanced customer experience and better labor management.

While the benefits of RPA (Figure 1) are similar to core platform transformation programs, RPA can deliver them much quicker, with lower risk, and at a fraction of the cost of traditional IT integration projects.

Figure 1: Benefits of RPA

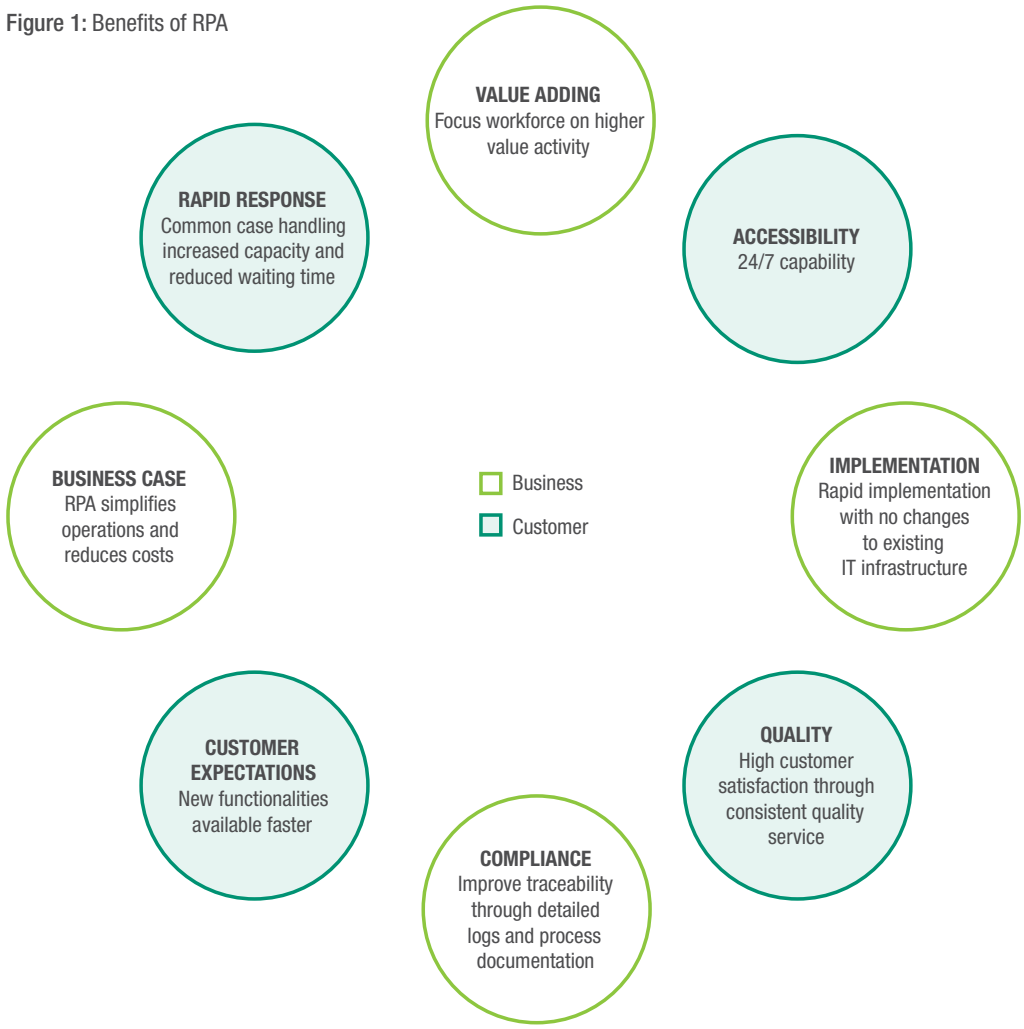
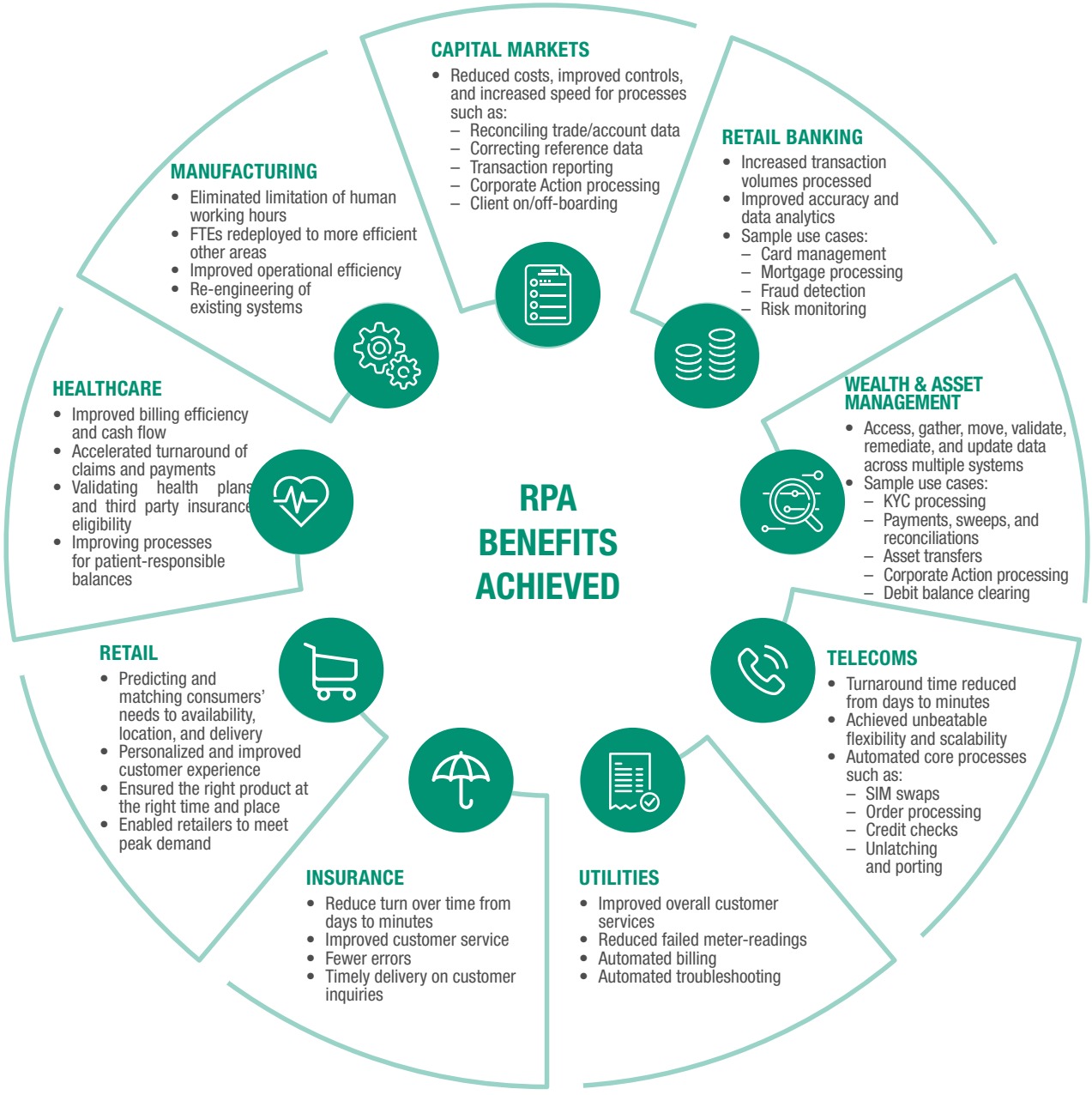


Figure 2: Benefits of RPA in different industries



Despite all of its benefits, however, it is imperative to understand that RPA will neither solve all automation challenges nor will it replace all existing applications in an organization. It is, nevertheless, an interesting option in various situations that organizations need to consider in order to remain relevant and competitive, as well as respond to changing customer expectations.

2.1 Comparing RPA within the financial services industry with other industries

Early RPA adopters have experienced significant increases in efficiency and productivity, in the range of two full-time equivalents (FTE) to as many as 20 FTEs for a single RPA implementation.

RPA is being adopted across a wide range of industries, such as manufacturing, healthcare, retail, energy, insurance, IT, telecommunications, and financial services, where we have observed roll-out of RPA implementations within multiple domains to enhance agility, accuracy, and compliance of service delivery (Figure 2).

In the telecommunications industry, Telefónica O2's RPA journey dates back to 2010, where RPA was initially used to automate 15 core business processes, which represented approximately 35% of all backoffice transactions in 2015. Telefónica O2 has managed to deploy over 160 robots across operations that process between 400,000 to 500,000 transactions per month. Over a three-year period, their investment in RPA has yielded an ROI of 650-800% and reduced the turnaround time from days to just minutes for some processes. Furthermore, Telefónica O2 has managed to achieve an unbeatable scalability, as their virtual workforce can be doubled almost instantaneously when new products are released and scaled back down after their introduction [Lacity et al. (2015)].

A large energy utility company is delivering around U.S.$6m in savings per year with RPA and another major telecommunications provider has estimated that for every U.S.$1 spent on RPA, it receives U.S.$8 in reduced operating expenditure and resources. Both companies have also achieved other benefits, including better overall customer service as a result of fewer errors and delivering on customer related inquiries in a timely manner [Grand View Research (2016)]. A large U.S. based manufacturer has been using RPA to improve operational efficiency as it proved to be an inexpensive and quicker alternative to reengineering the firm's existing systems. RPA has enabled the manufacturer to achieve 24-hour processing of payments, as processing is no longer limited by humans working in shifts, and enabled the manufacturer to redeploy 200 FTEs working on the order-taking process to other areas. The next step for the manufacturer is to enable RPA to assist the human workforce to perform highly labor-intensive work, such as credit checks by accessing, aggregating, and filtering data for them, in order to protect them from information overload [Tornbohm (2015)].

Some of the earliest and most aggressive RPA adopters are within the financial services industry, predominantly because of the major cost reduction challenges, regulatory and compliance pressures, and rigid legacy systems that they face. This has been driven by the potential benefits that RPA promises, as well as the fact that replacing a task with RPA can cost as little as one-third of the price paid to an offshore FTE and as little as one-fifth of the price for an onshore FTE [Chui et al. (2016)].

RPA is being employed within financial services organizations to better manage the increasing transaction volumes, move data for processing claims, card management (e.g., issuing replacements of stolen or lost cards), mortgage processing, as well as resubmission of failed payments. U.K.-based Co-operative Banking Group has automated over 130 processes with RPA, including complex CHAPs payment processing, VISA chargeback processing, audit reports, and other backoffice processes [Barnett (2015)]. The CHAPs payment process required on average 10 minutes to process one transfer request. The same request is processed within 20 seconds after applying automation, enabling the bank to maintain a bank-wide SLA on time allowed for CHAPs processing. The Co-operative Banking Group has achieved a number of benefits from implementing RPA in addition to the 80% savings in processing costs and the significant ROI for the bank with each process being deployed. For example, employees have been released to work on proactive customer account management, such as performing outbound customer calls every day of the week, enabling them to quickly identify customers in financial difficulty and proactively calling them to discuss their accounts.[2]

Another example from the financial services industry is Danske Bank, who have deployed RPA within their operations and backoffice functions on processes such as income payments and corporate actions processing. This has resulted in a 45% increase in employees' abilities to focus on customer related activities, 40% reduction on average process execution time, and elimination of human errors [Danske Bank (2016)]. Whilst many financial institutions are deploying RPA on operations and backoffice processes, RBC has taken a different approach by applying RPA across the trading

[2] This data was reported in a case study undertaken by Blueprism, entitled "The Co-operative Bank achieves 80 % saving in processing costs."

floor to improve efficiency and guarantee regulatory compliance [RPA & Artificial Intelligence (2016)].

Despite the potential benefits that RPA promises, many implementation attempts fail because RPA tools are not "plug and play" software and are often applied to broken processes, thus limiting their effectiveness. RPA tools require some degree of IT involvement to integrate with existing environments and buy-in from operations leaders to invest in the technology. Organizational change management and process transformation skills are also necessary to have in place in order to efficiently secure RPA deployment.

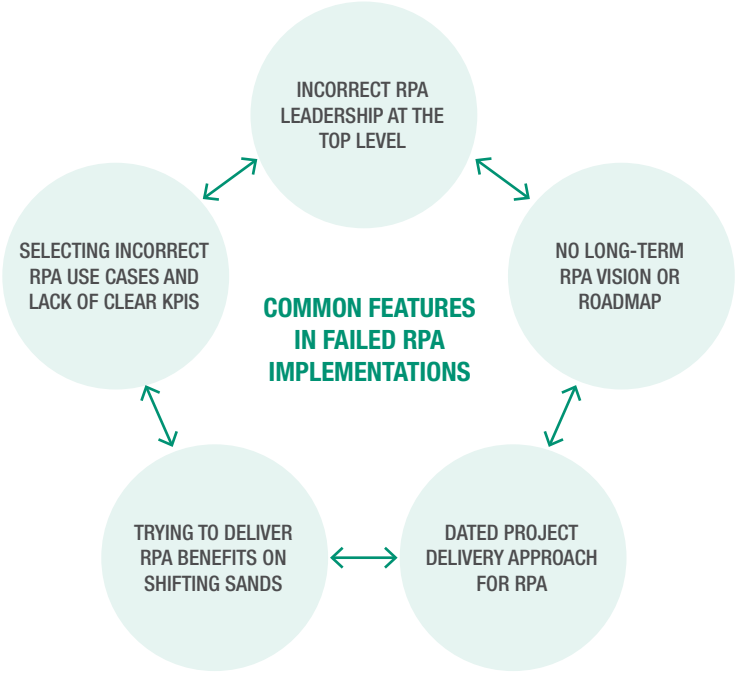
2.2 Why RPA is not living up to the hype

Early RPA adopters have managed to achieve significant economic benefits, but many more have run out of steam when trying to scale their initial pilot or proof of concept. In recent years, the hype of RPA has to some extent taken a hit as a result of strong negative opinions from individuals and organizations where RPA programs have failed to deliver the perceived benefits. This is a widespread problem not just for RPA but with emerging technologies in general. Media hype about the impact of robots in the human workforce does not help set realistic expectations. "Robots to steal 15 million British jobs in coming decades, warns Bank of England boss" was one of the headlines in newspapers across the UK in December 2016.

It is fair to say that "failed" RPA programs had issues far beyond the problems associated with how the technology was adopted. The challenges are at a much more fundamental level. The main issue is that RPA, without much due-diligence, is perceived as a silver-bullet to solve the three top challenges facing most businesses, namely cost reduction, efficiency gains, and acquisition and retention of customers, which it certainly cannot be. Furthermore, RPA is not the only platform to be used to overcome these challenges and deliver benefits. There are other ways and means to meet these challenges; the key to success is combining different technologies and key business decisions in change programs that span across people, processes, data, and tools.

RPA definitely has the advantage of providing lean and rapid benefits if deployed in an efficient way. However, some of the early adopters of RPA probably did not spend enough time to understand the "why" and jumped into "how" too quickly. Process automation programs in any organization start with an understanding of "as-is" processes, finding the gaps, and then agreeing the "to-be" state of these processes. RPA tools, however, are not designed to give benefits for processes that are immature, unstable, or broken. Robotic automation programs cannot deliver benefits to integrate with legacy systems if these outdated systems themselves are to be de-commissioned in the immediate future. The focal point of RPA use cases should not be the removal of human workforce; they should aim to improve accuracy, speed, agility, and remove the need for humans to execute repetitive tasks. Most failed RPA programs started without a clear definition on the financial, operational, and business KPIs. They were perceived as "another technology-led IT initiative" or were poorly structured without support from process change owners and key decision makers in business functions.

Figure 3: Common themes found in failed RPA projects



TOP 5 COMMON ISSUES FOUND IN OUR RESEARCH OF FAILED RPA PROJECTS AND PROGRAMS INCLUDE:

Issue 1: Incorrect RPA leadership at the top level

A successful RPA program has to be business led rather than IT led. All successful RPA projects have a common vital ingredient and it really is as simple as letting the business lead and use IT as a strong ally and partner.

The team should include IT infrastructure, IT security, architecture, risk and compliance, people or HR functions, finance, and all other key business functions in order to truly onboard the virtual workforce.

Challenges and recommendations:

Emerging technologies such as RPA, cognitive automation, and AI are often misunderstood to be a territory of the IT function within mid-size to large organizations. However, when it comes to RPA, it is key to remind ourselves that successful RPA programs aim to deliver benefits to the business and operations teams with a virtual workforce. Thus, the owners of business functions are best placed to lead the way and highlight the problem areas that can be tackled by technology enablement with RPA. Business teams are also fully empowered to understand which business processes would have the deepest impact and take proactive measures around human capital redeployment or downsizing. Business SMEs would also have a better understanding of their own processes, can easily train RPA robots, and play a pivotal role in measuring the outcomes of automation. Business operations heads should ultimately be accountable for defining the KPIs of selected RPA use cases. IT, on the other hand, should work with business stakeholders to follow a triage process in selecting the best use cases for early proofs of concept, building RPA technology infrastructure, and work in a collaborative way to lay the foundation for an RPA center of excellence. Joint governance between IT and the business is required to effectively manage RPA initiatives, make key decisions and remove obstacles.

Issue 2: Selecting incorrect RPA use cases and lack of clear business case, KPIs, or success criteria

Most failed RPA initiatives blame the insufficient outcome on technology. However it is the use case or candidate process that is typically one of the root causes. It becomes difficult for RPA initiatives to deliver the hyped benefits during a loosely defined RPA proof of concept (PoC) and answer the board level questions about strategic RPA sponsorship and funding without tangible benefits – usually cost saving and FTE reduction.

Challenges and recommendations:

There are still organizations that are either jumping straight into RPA vendor selection, or relying on IT or an RPA implementation partner to tell them which use cases to start with. These sets of events are usually driven by a PoC or pilot, but the long term strategic RPA benefits are often a mismatch. Instead of using the outcome of a PoC as a learning exercise for organizations, it is often put through scrutiny of skepticism. If the use cases identified for a PoC are lightweight, then it becomes hard to justify the RoI for the full-scale implementation. On the other hand, if the use-cases are complex, then it takes too long to get them right and test in a live environment to measure the benefits. This is where motivation levels drop and RPA initiatives are declared as over-hyped.

In reality, the business and technology stakeholders should work together to clearly define the business case, identify the real drivers in the organization for adopting RPA, and define use cases with clear and tangible outcomes and KPIs. Project sponsors and senior stakeholders should be responsible for signing off the PoC use cases and pre-agreed success factors. RPA PoCs should be seen as a learning exercise with a feedback process in place to understand, improve, and evolve for next time.

Issue 3: No long-term RPA vision or roadmap:

Most organizations in the early stages of automation adoption have no RPA experience. This leads to knee jerk reactions when it comes to making strategic decisions around RPA. Organizations are keen to explore in order to realize the benefits of RPA but often lack the conviction and vision to set up a long-term direction. There are pockets of supporters of RPA but also equally skeptical individuals.

Challenges and recommendations:

Organizations with successful RPA initiatives have a strategic vision that is usually achieved through the setup of an RPA center of excellence and a strong governance structure. In addition to utilizing industry experts (normally from RPA vendors or implementation partners), there is an equal focus and commitment to nominate internal IT and business representatives to drive the RPA strategy. This replaces skepticism with constructive feedback, creates opportunities for organizations to learn RPA, and fully understand its merits and limitations. Furthermore, corridor conversations, such as “I heard RPA failed to deliver any value in my previous company, are we sure we want to do it?“ are avoided. For successful RPA initiatives, organizations have to seriously commit and be ready to get their hands dirty. If organizations start with a skeptical view, and hence limit their involvement and

commitment, the outcome is bound to be unfavorable. RPA is a proven concept, but it needs to be carefully set up within any organization that is new to the technology.

Another issue is that once the robotic automation has replaced human intervention, the staff are quickly mobilized into doing other work or exception handling. For certain cases, this makes sense but without a clear mandate between HR and the heads of business, it is unlikely the savings from human FTEs versus virtual workforce can be achieved as the operational cost remains static and in fact increases as the business also has to pay for the RPA investment.

Finally, another pitfall is getting over-ambitious and trying to achieve too much process automation with a large RPA transformation program. There are other aspects and methods of automating processes that should not be forgotten. When it comes to complex automation use cases, which require fixing data sources, exception handling, and sometimes even automating broken business processes, it is well advised that one should take a step back and look into a more holistic view of end-to-end process change.

Issue 4: Trying to deliver RPA benefits on shifting sands

Organizations often try to reap RPA benefits on "shifting sands." This is true in organizations on a transformational journey where the people structure, business processes, and underlying tools and technologies are constantly changing. RPA use cases are not best suited to deliver benefits when the processes are not mature or there could be alternative treatment to these changes for far better benefits.

Challenges and recommendations:

RPA is not a silver bullet to solve every type of automation problem or achieving the perfect 24/7 virtual workforce. Even mature and industry leading organizations sometimes make the mistake of embarking on an RPA journey based on use cases that are fundamentally going to change. Examples include trying to automate the swivel chair problem of legacy applications not communicating with each other, whilst there is a conflicting IT portfolio rationalization program under strict NDA trying to decommission these legacy applications.

Issue 5: Dated project delivery approaches for RPA

A common problem in most organizations is imposing heavy IT project delivery methodologies on RPA projects, with often time consuming, low-value documentation, management packs and management information reports being produced.

Such bureaucracy erodes one of the key advantages of RPA: rapid development and deployment into production. Typically, process automation with RPA takes no more than 2-4 weeks from inception to production. The traditional waterfall methodology cannot keep up with the pace of RPA delivery. As such, it is paramount for organizations that are looking to adopt RPA to review their delivery approaches and adopt agile delivery methodologies. Organizations that adopt agile and lean delivery methodologies when embarking on an RPA journey have much higher success rates than those that follow traditional delivery methods.

3. HOW IS THE FUTURE OF THE VIRTUAL WORKFORCE GOING TO UNFOLD?

3.1. The untapped potential of RPA

RPA seeks to automate business tasks that are mundane, highly repetitive, rule-based, and use structured data to make them more repeatable, faster, and less prone to human errors. Business processes and tasks that are most suitable for RPA are the ones that involve a considerable amount of repetitive key strokes. This means industries that have numerous manual interventions, rapid hiring cycles, and suites of software applications have potential of at least some, if not major, process automation deployments [Tornbohm (2016b)].

RPA can be used to read data on the aforementioned systems, act as an intelligent web data extraction tool, manipulate data, and subsequently input the outcome to other systems for different processes, yet remain simple in its usability [Greer and Beattie (2016)]. The use of RPA in similar scenarios is more robust than using a traditional screen-scraping technology, as RPA is capable of supporting changes to data fields and can be deployed on an enterprise level [Tornbohm (2016a)].

RPA in financial services is estimated to carry out tasks much faster while performing the work of three FTEs at a third of the cost; which means that financial institutions are able to increase their efficiency gains by a multiple of nine [Chui et al. (2016)]. Other industries that have

similar cost saving and waste reduction pressures from trade or industrial bodies, governments, and other stakeholders represent the currently untapped potential of RPA. Furthermore, RPA use-cases in current industries and new market solutions are expected to increase in the coming years.

3.2 Long term impact of RPA

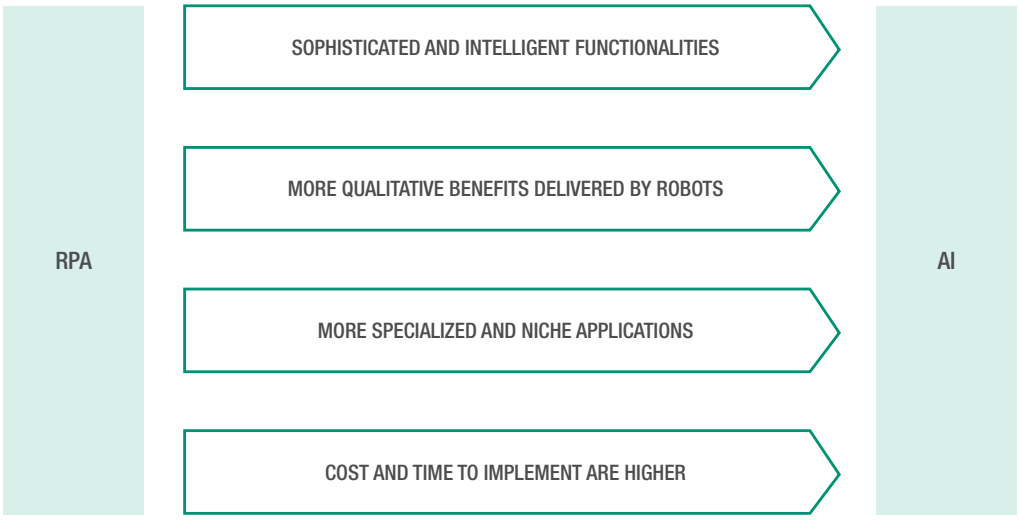
RPA and AI will impregnate a wide segment of our daily life in the next decade, with huge implications across various industries. However, as much as the predictions for the evolution of technology are largely consistent, some opinions are deeply divided on how advances in RPA and AI will impact the economic and employment picture over the next decade. Some have painted a future in which significant numbers of both blue and white-collar jobs are destroyed by automation. Many experts are expressing concerns that greater computerization of the workplace will lead to increases in income inequality, unemployment, and disruption of the conventional social structure. On the contrary, many expect that technology will not take away more jobs than it creates in the next decade. In fact, since the dawn of the Industrial Revolution, technology has been killing and generating jobs, and has benefited mankind in many ways, some of which are now taken for granted in a variety of industries.

3.3. Continuing the automation journey

In the ongoing marriage of mind and machine, the neuro-physiological portal through which a fuller integration could occur was opened marginally two decades ago by some promising theoretical conceptions. The present generation is rapidly passing through that conceptual window. However, this race for a greater human-machine intimacy may be more than simply another step along the unique road of history. If the present vector of self-destructive progress continues, it may be that this avenue of development is the one that holds the greatest (some would say, only) promise for salvation [Hancock et al. (2013)].

Future customer engagement models will work from our fingertips through social relationships with organizations; enabling quicker access to different services at any time and with better quality. For this purpose, “cognitive process automation” (CPA) widens the application of RPA to more knowledge-based work, such as extracting information from unstructured

Figure 4: Progression in characteristics as we climb the generations of robots from traditional RPA to AI



sources and enhancing decision-making. Cognitive agents have a self-learning capability that enables them to act and learn from experience, from humans, and even on their own, thereby developing the ability to interact with their own environment. CPAs can help with work that requires judgement and perception, enabling RPA to reach a new level. Combining RPA with cognitive agents provides a more strategic perspective that has the power and potential to deliver business results, such as greater customer satisfaction and increased revenues by going above and beyond basic RPA. It is no surprise then for financial institutions to have CPAs (like chatbots, machine learning, AI, natural language processing (NLP), speech recognition, etc.) on their minds when looking at reinventing the customer experience, whilst also cutting costs in roles that are ripe for automation. For example:

Chatbots can communicate through several channels, such as messaging apps (e.g., Slack, Facebook), SMS, text, or voice-based assistants (e.g., Siri, Alexa).

- Machine learning can make predictions about process outcomes by identifying patterns and prioritizing actions depending on predicted outcomes.
- NLP, speech, and image recognition can facilitate understanding of free flow sentences and convert speech audio, text, or images into structured information.

Combining the above with RPA would enable robots to learn from their experience of process execution, enabling them to handle exceptions, manage unstructured data, and actually improve over time. Thus, cognitive RPA can be used to support employees and customers over phone or via chat, such as in employee service centers. A U.K. auto insurer saw a 22% increase in conversion rates, a 40% reduction in validation errors, and a 330% overall RoI following the implementation of such cognitive technology.

As we progress from traditional RPA to AI, we will observe several evolutions in characteristics (Figure 4).

The adoption rates and ability to use machine learning and AI will continue to shake up the outsourcing world in the years to come. The risk of human error is always going to be high and by employing a central AI function some of the risk is mitigated.

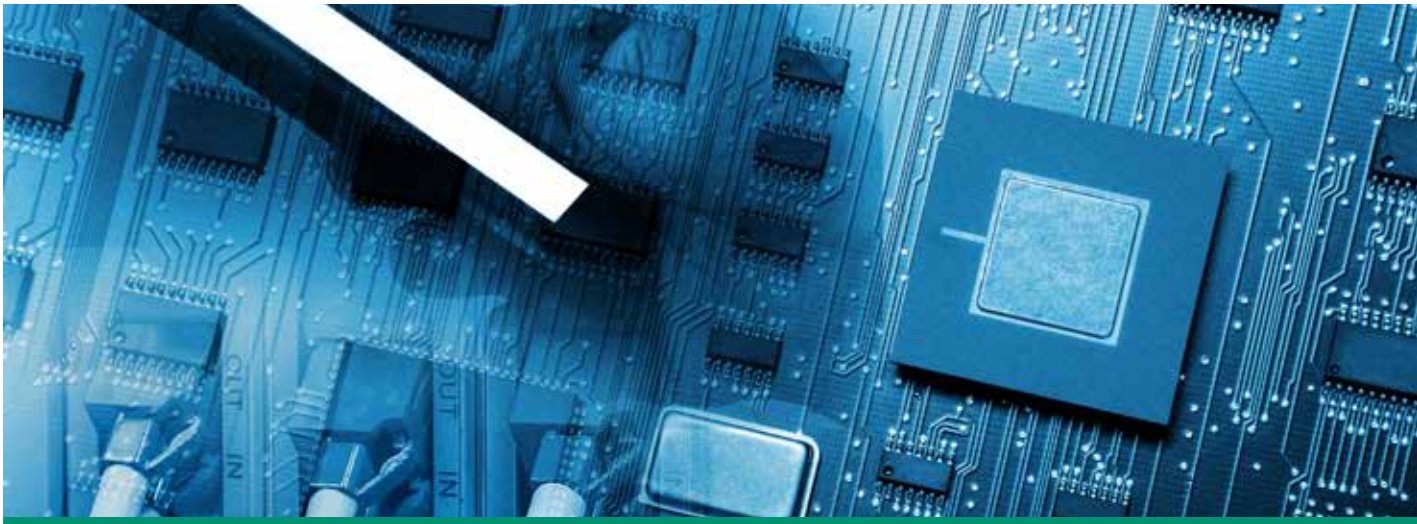
CONCLUSION

As organizations continue to explore and expand their use of new technologies to solve the top three challenges facing most businesses, namely cost reduction, efficiency gains, and acquisition and retention of customers, the need for humans to interact and collaborate with robots will increase, thus redefining the required capabilities of the future workforce.

Understanding which technology to deploy, where, and how is challenging. Organizations need to understand the differences between the plethora of automation

and cognitive tools that have machine learning or AI capabilities, where these types of tools have been deployed, and how they will likely evolve in the near future. Furthermore, how to combine and deploy them into an organization's unique IT and process landscape poses a major challenge, as these tools are not "plug and play" and organizations have poor insight into, and knowledge of, the "where" and "how" to use them.

Even though RPA holds high potential of fully eradicating most of the manual and repetitive tasks performed by humans, RPA tools need to be evaluated against other automation and cognitive tools. Organizations should use a structured approach in identifying and cataloguing unautomated processes in order to determine which are most suitable for RPA or other automation or cognitive tools and understand how these tools can support various key business initiatives.



References

Barnett, G., 2015, "Robotic process automation: adding to the process transformation toolkit," OVUM

Chui, M., J. Manyika, and M. Miremadi, 2016, "Where machines could replace humans – and where they can't (yet)," McKinsey Quarterly, June

Danske Bank, "Robotic Process Automation (RPA) within Danske Bank"

Grand View Research , 2016, "Robotic process automation (RPA) market size report, 2024," http://bit.ly/2jfxlo2

Greer, S., and C. Beattie, 2016, "Robo op : how robotic process automation is applied in banking operations," Celent, http://bit.ly/2wNThzb

Hancock, P. A., R. J. Jagacinski, R. Parasuraman, C. D. Wickens, G. F. Wilson, and D. B. Kaber, 2013, "Human-automation interaction research: past, present, and future," Ergonomics in Design: The Quarterly of Human Factors Applications., April 23,

Infosys Consulting, 2017, "Robotic process automation," http://gtnr.it/2xaq6bS

Lacity, M., L. Willcocks, and A. Craig, 2015, "Robotic process automation: mature capabilities in the energy sector," Outsourcing Unit Work Research Paper Series, no 15/06, London School of Economics and Political Science

Mckinsey, 2017, "A future that works: automation, employment, and productivity," Mckinsey Global Institute

Tornbohm, C., 2015, "Gartner use cases for robotic process automation: providing a team of 'virtual workers,'" Gartner, October

Tornbohm, C., 2016a, "Robotic process automation: eight guidelines for effective results," Gartner, October, http://gtnr.it/2xaq6bS

Tornbohm, C., 2016b, "Market guide for robotic process automation software," Gartner, November

Vishnu, S., V. Agochiya, and R. Palkar, 2017, "Data-centered dependencies and opportunities for robotics process automation in banking," Journal of Financial Transformation 45, 68-76

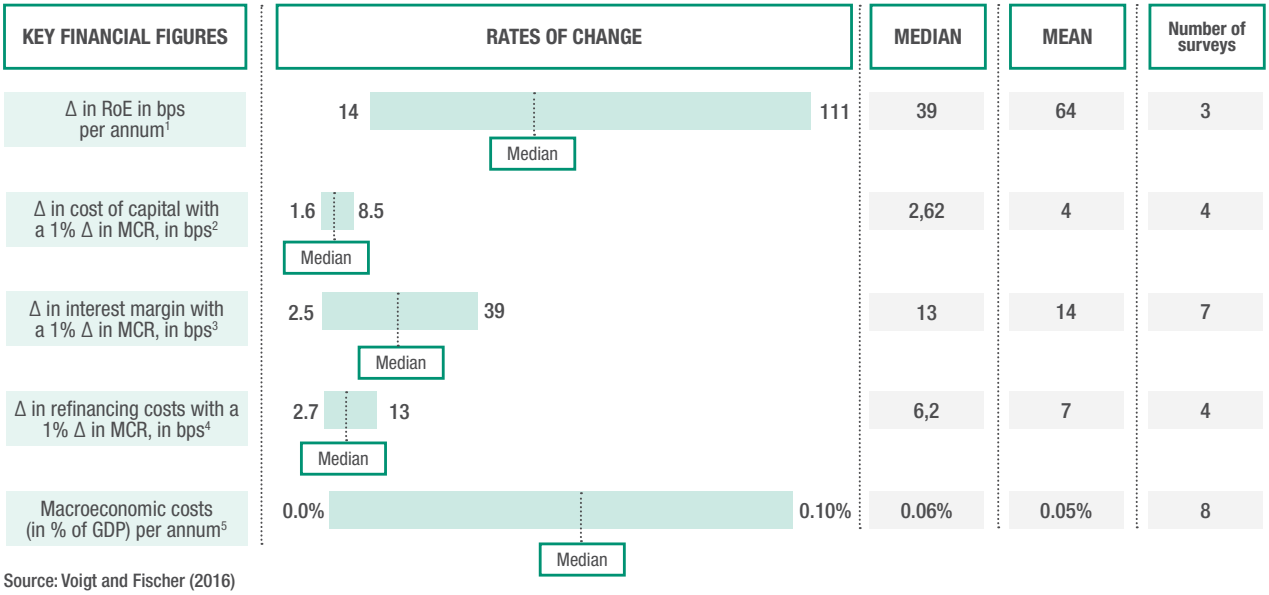
The impact of financial regulation on business models of cooperative banks in Germany

MATTHIAS FISCHER | Professor of Banking and Finance, Technische Hochschule Nürnberg Georg Simon Ohm, Germany; Adjunct Professor of Banking and Finance at IAE Université Nice Sophia Antipolis, France

ABSTRACT

A number of studies have highlighted the potential negative implications of stronger financial regulations, however, only a few studies have attempted to quantify the regulatory impact of Basel III on profitability. Regulation has specific costs, as well as benefits, for any economy. Likewise, it has consequences for the cost of capital of banks, as well as their interest margin. The analysis provided in this article has calculated the implications of Basel III on the profitability of banks and found that they range between 14 and 111 basis points – in case no countermeasures are taken by the respective banks. In addition, this article looks at the implications of interest rate risks on banks, and the potential negative impact on bank capital ratios in the case of interest rate risk integrated in the capital requirements of pillar 1 of Basel III. Consequently, using the balance sheet data from 756 cooperative banks in Germany, we have examined the implications of the "Basel interest rate shock," where a sudden parallel shift in the yield curve of 200 basis points happens. The three test calculation scenarios assume the following: (1) a full implementation of Basel III without the integration of interest rate risks in the banking book of pillar 1, (2) analysis of theoretical maturities for the calculation of the interest rate risk, and (3) using legal contract terms and maturities as the basis for calculation of the interest rate risks. The results of the study show that in a scenario where the legal contract term was used, 5.3% of the analyzed group did not reach the minimum ratio for core capital of 4.5%, and another 46.6% of the banks would be below the 7% ratio and, therefore, would be limited in their earnings distributions; 86.9% of the cooperative banks in the analyzed group would fall below the threshold of 10.5%. We reach the conclusion that financial regulation should not follow the rule of "one-size-fits -all" because the business models of small cooperative banks in Germany are different to those of major global banks. A global or European uniform regulation for all banks, neglecting size and business model, could also jeopardize the culture of fixed interest financing for mid- and long-term loans for German SMEs.

Figure 1: Implications of regulatory changes on key financial figures



Source: Voigt and Fischer (2016)
[1]Yearly change in return on equity (ROE) in basis points in case no countermeasures are induced.
[2]Change in weighted average cost of capital (WACC) in basis points if minimum capital requirements (MCR) change by one percentage point.
[3]Change in interest margin in basis points (spread between deposit rates and credit rates) in case MCRs change by one percentage point.
[4]Change in cost of borrowing capital in basis points in case MCRs change by one percentage point.
[5]Cost for the respective economy, including economic benefits of regulation, stated as percentage of GDP per year.

1. THE IMPACT OF REGULATION ON THE PROFITABILITY OF CREDIT COOPERATIVES[1]

Since January 1, 2014, the Basel III regulation has been implemented in Europe under the auspices of the Capital Requirements Directive IV (CRD-IV). The key requirements of the new regulation are: a more stringent definition of regulatory capital, greater weighting for core capital, higher minimum capital ratios, the introduction of an anticyclical buffer as well as a leverage ratio, stricter requirements for liquidity [liquidity coverage ratio (LCR) and net stable funding ratio (NSFR)], and a more significant consideration for counterparty risks.[2]

There are numerous studies that look at the negative consequences of the more stringent regulations on the profitability of banks, but only a few quantifications are available for the relevant financial ratios. Consequently, we examine the existing regulatory studies to derive profitability parameters for the forecast calculations in banks.

Banks can use different strategy alternatives to respond to regulatory changes. In the studies analyzed, the alternative strategies are simulated either as an ad-hoc measure or as an optimization measure over the course of time. Many studies focused predominantly on the consequences of the new minimum capital requirements.

Overall, the results of 23 studies have been examined; however, only 13 provide comparable results to allow for the estimation of delta parameters for the impact of regulation. In terms of applying delta parameters, the studies need to be adapted according to the size of the credit institutions, their business models, and their respective countries.

The various studies also have different assumptions regarding growth for core capital [common equity tier 1 (CET1)], the additional core capital (Tier 1), and the supplementary capital (Tier 2). Most studies refer to CET1 capital and additional Tier 1 capital, while Tier 2 capital is not included in their calculations.[3] In empirical studies, the analyzed banks are mostly larger institutions of different countries, whereas the rather small credit cooperatives have hardly been considered. Table A1, in the Appendix, presents the studies considered.

[1] Results of the analysis have been published in Voigt and Fischer (2016).
[2] Additional capital will be required for systematically important financial institutions (SIFIs)
[3] Exceptions are the studies by McKinsey (2010) and BCG (2011), where a full implementation of the new regulation has been assumed at the time of the study.

Figure 1 shows the estimated annual negative impact of regulations on the return on equity (RoE) until the full implementation of the capital requirements in 2019. The changes range from Delta-RoE minus 14 basis points (bps) to minus 111 bps, with the median at minus 39 bps and the mean value at minus 64bps. However, there are large differences in the assessment criteria used in the RoE calculations. For example, the NSFR effect is explicitly taken into account in one study, while in another study the return of EBT is used instead of RoE. Accordingly, a team knowledgeable in banking has to develop adaptations of the delta parameters in Figure 1 when applying the spreads for bank profit forecast calculations.

2. INTEREST RATE RISK IN BANKS AND REGULATORY CAPITAL REQUIREMENTS

Management of interest rate risks is traditionally a key component of the business model of banks. Credit institutions have the task of converting short-term deposits into long-term loans, also referred to as term transformation.

A rising interest rate curve is a prerequisite for the generation of income from the term transformation. In Germany, interest rate curves over several years – the average interest rate structure – are typically upwards sloping.

Figure 2 demonstrates that the shapes of the interest rate curves are by no means static. The interest yield curves for the U.S. are shown for the period between September 2007 and September 2015. The yield is exposed to permanent interest rate fluctuations in the various maturities. A parallel shift in the overall interest yield curve is not the rule. Rather, changes in interest rates also occur independently of one another during the individual maturities and ensure varied yield curves of treasury bonds. In the years 2007 to 2015, the U.S. yield curve has maintained its normal structure. The challenge for credit institutions is primarily to correctly forecast future changes in the interest rate structures and to implement appropriate management measures while at the same time meeting regulatory requirements.

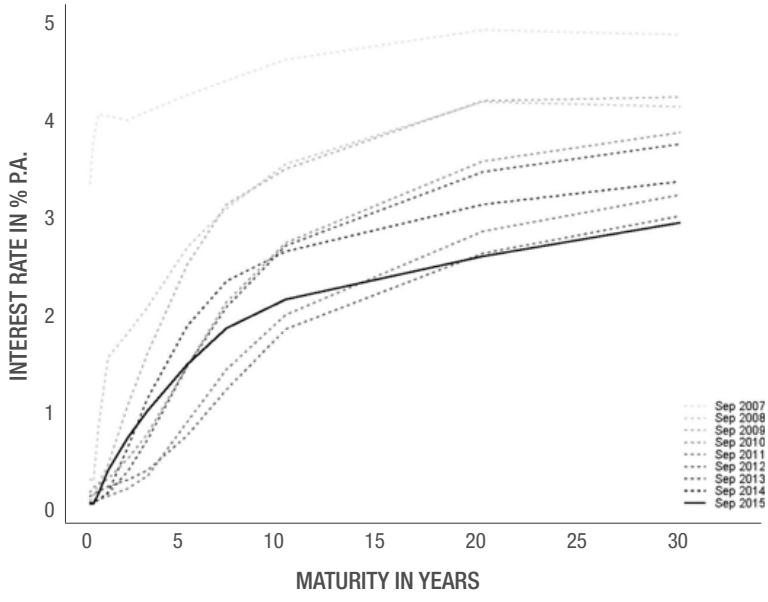
From an accounting perspective, interest rate risks in banks can be incorporated with both the trading book and the banking book. By the end of 2013, the distinction between trading book and banking book was regulated in section 1a KWG Kreditwesengesetz. Since 2014, the distinction has been anchored in Regulation (EU) No. 575/2013 of the European Parliament and of the Council, the so-called Capital Requirements Regulation (CRR). However, no significant changes were made with regards to the allocation criteria for the trading or banking book [Weigel and Sierleja (2015)].

Transactions for the purpose of generating a profit for the bank by the short-term use of existing or expected differences between bid and ask prices or the utilization of market price fluctuations are part of the trading book. The banking book is a residual value and covers all transactions that cannot be assigned to the trading book.

Interest rate risks in the banking book have so far not been part of the quantitative minimum capital requirements according to pillar 1 of the Basel framework, but have been allocated in the "qualitative" pillar 2. Pillar 2 requires appropriate risk control and controlling processes for interest rate risks in the banking book. All the risks identified in Pillar 2 of the Basel Regulatory Approach must also be included in the calculation of the risk-bearing capacity of the banks.

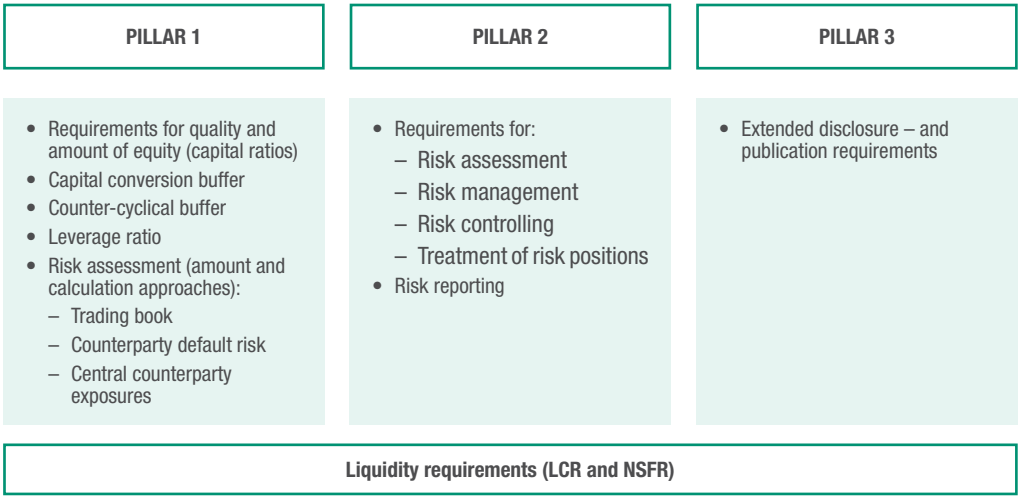
Pillar 1 of the Basel rules aims to standardize the capital requirements in order to ensure their international comparability. In pillar 2, institutions are granted more freedom with regard to the individual capital and risk assessment. Financial institutes should identify all risks of relevance, quantify them with suitable methods, and

Figure 2: U.S. Treasury Yield Curves per September for the years 2007 until 2015



Source: Voigt and Fischer (2016), with data from U.S. Department of the Treasury, 2015

Figure 3: The pillars of Basel II with specific consideration of Basel III



Source: Voigt and Fischer (2016)

provide them with adequate capital. The reasons for this kind of differentiation between pillar 1 and 2 are the different risk circumstances of individual banks, which could not justify a complete standardization. In addition, the two pillars differ in the fact that the first pillar focuses on the current business, while the second pillar also considers future changes in the bank's own business or the market environment [Deutsche Bundesbank (2013)]. Figure 3 summarizes the most important requirements of the three-pillar model.

3. INTEREST RATE RISK STRESS-TEST AND CALCULATION METHODOLOGY

The Basel Committee on Banking Supervision is aware of the interest rate risk in the banking book. However, the originally planned integration of the interest rate risks into the first pillar of Basel II was not realized due to a lack of an international consensus regarding the calculation methods for interest rate risks. There is no international standard procedure for the parameterization of asset and liability positions with undefined capital or interest rate fixation. In addition, with regards to a periodic or present value approach of interest rate risks there is no common standard used by supervisory bodies in the various countries, as national market structures for credit institutions are also very different [Deutsche Bundesbank (2012), Basler Ausschuss für Bankenaufsicht (2006), Österreichische Nationalbank (2008), BaFin (2014.)

As part of the implementation of Basel III, the capital adequacy for interest rate risks in the banking book is discussed in order to limit the bank's risk of insolvency. Additional equity should serve as a risk buffer for losses from unexpected changes in market interest rates. The supervisor examines the implications of an ad-hoc increase or fall in market interest rates by 200 basis points [Fischer and Heil (2015a)]. This indicator is also known as the "Basel interest rate shock" and measures the present value effects of an unexpected interest rate change on the company's own funds. It has to be determined in accordance with BaFin circular 11/2011 and must be reported to the supervisory authorities, but financial institutions are allowed to choose the yield curve used in the internal calculations. All banking book positions with undefined periods for capital commitment and interest rate fixation, or with implicit interest rate options, must be adequately represented in the internal calculations of the bank. However, the methods and procedures for the calculation of interest rate risk must meet the minimum requirements for risk management (MaRisk). Credit institutions with a negative change of more than 20% in the present value of the regulatory capital are classified as being of higher interest rate risk [BaFin (2011)].

The Basel interest rate shock only considers the present value effects of a sudden parallel shift in the yield curve. The net present value calculation helps to improve the comparability between credit institutions, but is particularly controversial in the case of banks that use the P&L-oriented methodologies to monitor

their interest rate risks. For credit institutions with P&L guidelines for interest rate risk management, the German supervisory body, BaFin, offers an alternative procedure for estimating the change in present value; the potential disadvantage of the alternative calculation method could be an overestimation of the risks by using specified modified durations for the respective maturities in contrast to using internal models [BaFin (2011)].

A standardized calculation methodology for interest rate risk would be preferred for international comparisons of its implications. A regulatory model that avoids undesirable effects for banks and their clients should consider incorporate the following [Fischer and Heil (2015a)]:

- A standardized composition of the financial ratio.
- A dynamic interest rate scenario specified by the supervisor.
- The capital requirement for market value and interest rate risks in the banking book.
- The consideration of individual business models of credit institutions or the definition of a threshold as the upper limit for an interest rate risk that is not subject to capital requirements.

4. SIMULATION FOR THE CALCULATION OF INTEREST RATE RISK AND NEW CAPITAL REQUIREMENT RULES

4.1 Database and assumptions

The impact of the possible capital requirement for interest rate risks was tested in the year 2015 based on data from 756 cooperative banks in Germany [Fischer and Heil (2015a, b)]. For the legal duration of the respective financial contract, the following calculation assumes an approximation for the change in the present value; the present value is calculated on the key date and no further possible balance sheet or profit growth is taken into account. The calculation considers the present value calculation of interest rate risks for the banking book and a static position of the bank's capital. The simulation assumes an immediate implementation of CRD IV regulation. The equity capital employed in the simulation was adjusted in accordance with the CRR Capital Requirements Directive. The risk weighted assets were increased by 1.4% for the Basel III scenario.[4]

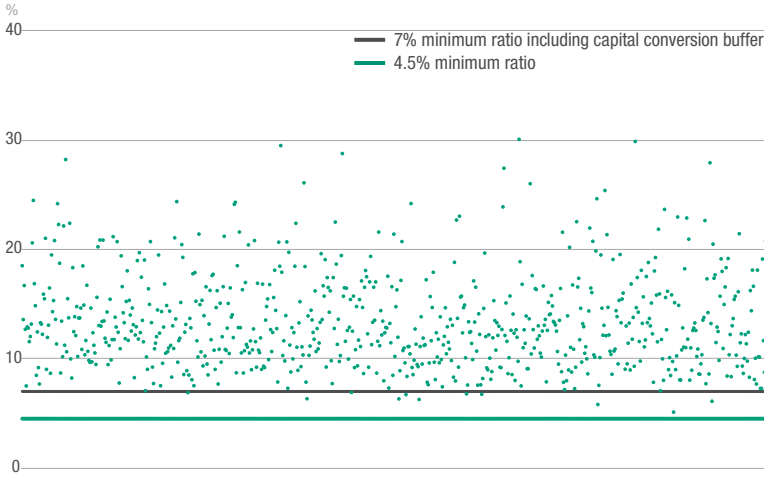
4.2 Three interest rate scenarios for the core capital ratio

Three scenarios are presented for the effects of the regulatory changes on the core capital ratio of 4.5% and of 7%, including the capital conservation buffer. Figures 4, 5, and 6 show the three scenarios of regulatory capital adequacy for interest rate risks with respect to the core capital ratios. Scenario 1 involves the core capital ratios of the analyzed banks for the adjustment to Basel III without the integration of interest rate risks in the banking book in pillar 1 of Basel III. Scenario 2 involves the core capital ratio with the assumption of theoretical or fictitious maturities for the calculation of the interest rate risk. Scenario 3 includes the extreme scenario of the legal contract term and maturities as the basis for calculation of the interest rate risks.

Figure 4 presents the results of scenario 1, examining the core capital ratios of all banks exclusively on the basis of the CRR and CRD IV requirements and excluding the interest rate risks in the banking book. All of the banks remain above the minimum ratio of 4.5% and only 1.5% of the banks remain below the minimum ratio of 7% (4.5% plus 2.5% capital retention buffer). Only 1.5% of the banks would be sanctioned with restrictions on the payout of earnings [Fischer and Heil (2015a), Voigt and Fischer (2016)].

Figure 5 presents the results of scenario 2, calculating the core capital ratio according to CRR as well as the

Figure 4: Core capital ratio according to CRR without integration of interest rate risk in pillar 1 – scenario 1



[4] The increase of 1.4% was based on estimations of Deutsche Bundesbank (2015) and Fischer and Heil (2015a).

interest rate risks being subject to capital requirements. The assumption of theoretical maturities was used to calculate the maximum present value loss due to interest rate risks. As a result, the equity ratio deteriorated significantly. Only 0.7% of the banks did not reach the minimum ratio of 4.5% for the core capital and have to adapt their business models immediately or create ad-hoc additional common equity. A total of 20.8% (0.7% + 20.1%) of the banks in the analyzed group would have to limit their earnings payout because they are below the hurdle of 7%.

Figure 6 presents the results of scenario 3, where the legal contract term of the balance sheet items is used as the basis for the interest risk calculation instead of the theoretical maturity. 5.3%, or 40, banks of the analyzed group do not reach the minimum ratio for core capital of 4.5% and another 46.6% of the banks would be below the 7% ratio and, therefore, would be limited in their earnings distributions.

4.3. Three scenarios for the regulatory equity ratio and interest rate risk

In the next step, three regulatory escalation steps are simulated for the minimum capital ratio. Figures 7, 8, and 9 illustrate the impact of the integration of interest rate risks on regulatory capital ratios of 8% and 10.5%, respectively, including capital conservation buffers.

Scenario 1 examines the capital adequacy ratios of the analyzed banks for the adjustment to Basel III but without integration in Pillar 1. Scenario 2 looks at the capital ratios for the calculation with theoretical maturities. Scenario 3 examines the extreme scenario of the interest rate calculation with legal contract terms for all balance sheet items.

Figure 7 presents the results for scenario 1, where the regulatory capital ratios of all banks are calculated on the basis of the requirements according to CRD IV, without taking into account the interest rate risks in pillar 1. Overall 17.9% (2% + 15.9%) of the banks are below the minimum ratio of 10.5% (8% plus capital conservation buffer of 2.5%), 2% are below the hurdle of 8%, and a further 15.9% must be subject to earnings distribution restrictions.

Figure 8 presents the results of scenario 2, and shows the integration of the interest rate risks in pillar 1 with the assumption of theoretical maturities for balance sheet items. 20% of the banks (58 banks) do not reach the minimum ratio of 8% and another 40.4% of the

Figure 5: Core capital ratio according to CRR with integration of interest rate risks in pillar 1 – scenario 2

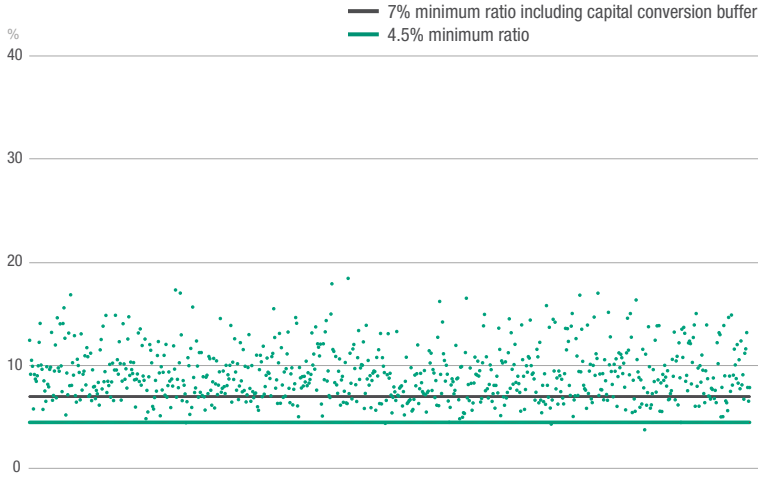
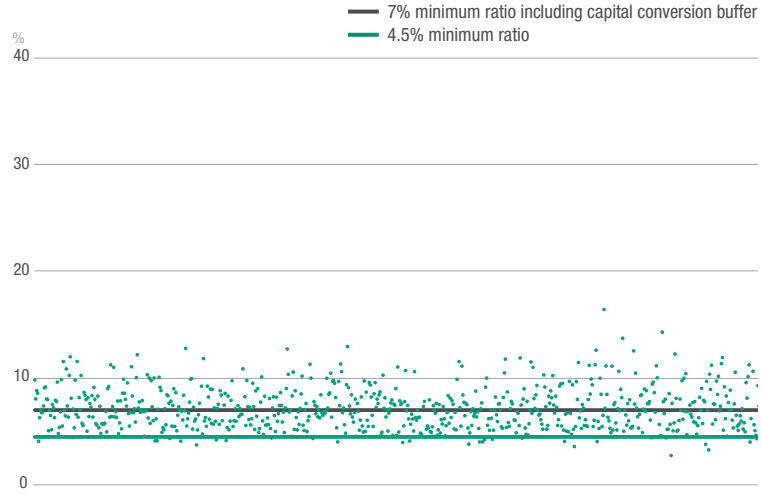


Figure 6: Core capital ratio according CRR with integration of interest rate risk in pillar 1 – scenario 3



Sources: Fischer and Heil (2015a) and Voigt and Fischer (2016)
Database: 756 German cooperative banks from the year 2013; the analysis is reporting date related – no budgeted balance sheet, no earnings growth considered. The change in present value due to interest rate risks is an approximation.

analyzed group would have to restrict their earnings payout as they are below the minimum ratio of 10.5%, including the capital conservation buffer.

Figure 9 presents the results of scenario 3, and applies the extreme scenario of the legal contract term as calculation basis for interest rate risks. A total of 81 banks, or 53.3%, are below the minimum capital ratio of 8%; a further 33.6% are below the minimum ratio of 10.5% including capital conservation buffer and would thus be restricted in their earnings payout potential. In the extreme scenario of the legal term of the contract for all balance sheet items, a total of 86.9% of the banks under investigation would be below the threshold of 10.5%.

5. TERM TRANSFORMATION AND INTEREST RATE INCOME

In a sustained low-interest rate environment, Deutsche Bundesbank sees the risk that financial institutions with low profitability will be open to take more risks and that they will try to compensate the lower interest income with a higher structural contribution; this refers in particular to savings banks and credit cooperatives, which are strongly dependent on the net interest income [Deutsche Bundesbank, (2014)].

The interest contribution is calculated as the difference between the agreed customer interest and the interest income from a fixed-term investment in the money and capital market with the respective maturities. The structural contribution is mainly the result of different maturities of interest rates regarding assets and liabilities generated by term transformation [Becker and Peppmeier (2011)].

The structural contribution has a significant impact on the interest income of savings banks and credit cooperatives. According to Memmel (2010), bank-specific management decisions are responsible for 83% of the adjustments of interest rate risk. In contrast, the regulatory quantitative limitations of interest rate risk in Basel II is only relevant for 8% of the changes. Table 1 presents the respective proportion of the interest income resulting from term transformation for

Sources: Fischer and Heil (2015a) and Voigt and Fischer (2016)
Database: 756 German cooperative banks from the year 2013; the analysis is reporting date related – no budgeted balance sheet, no earnings growth considered. The change in present value due to interest rate risks is an approximation.

Figure 7: Regulatory capital ratio according to CRR with integration of interest rate risk in pillar 1 – scenario 1

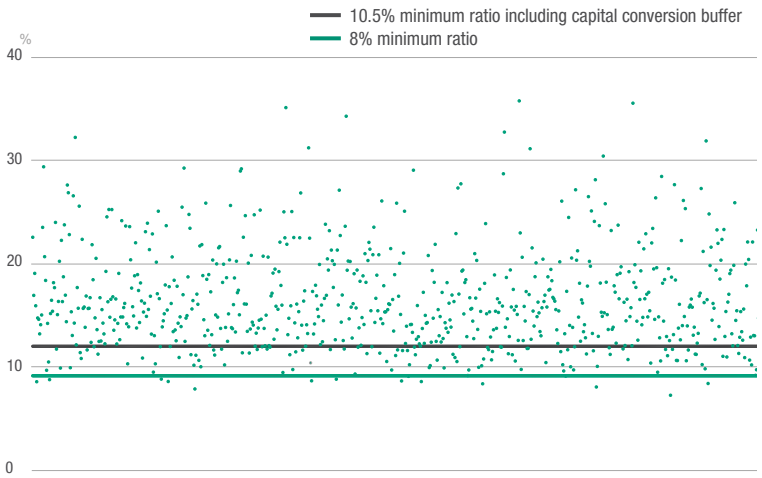


Figure 8: Regulatory capital ratio according CRR with integration of interest rate risk in pillar 1 – scenario 2

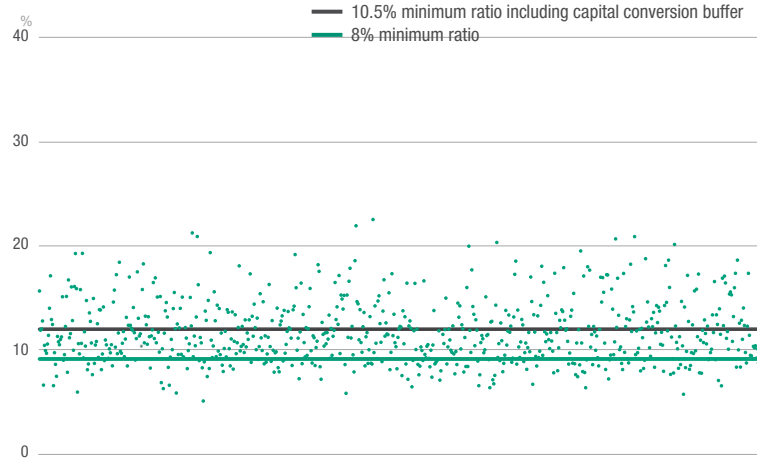
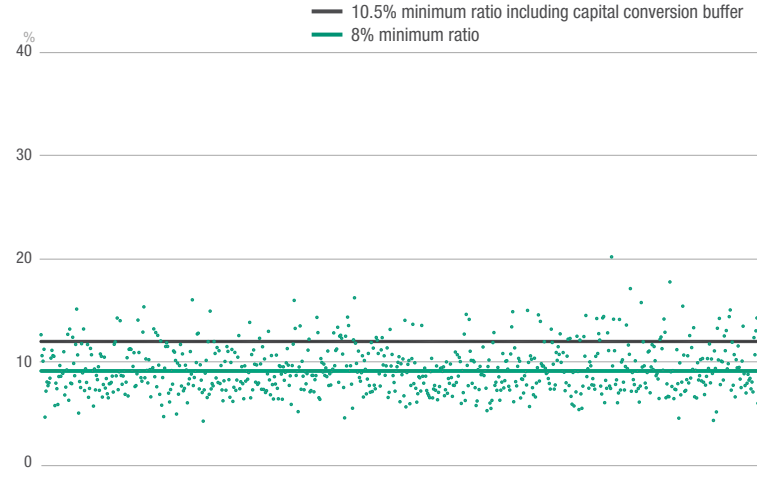


Figure 9: Regulatory capital ratios according CRR with integration of interest rate risk in pillar 1 – scenario 3



different banking groups, and is subject to significant fluctuations over time. For credit cooperatives, the proportion of interest income resulting from term transformation is 4.7% in 2008 and 24.8% in 2009.

6. INTEREST RATE RISK AND LONG-TERM FINANCING HABITS

Interest rate risks in the banking book are a major risk type as well as an important source of income for many banks. Term transformation also has macroeconomic implications, since it matches the different consumption and investment patterns of individuals and companies. In the case of long-term financing in Germany, fixed rates provide planning certainty for small- and medium-sized enterprises (SMEs), as well as for private individuals. A change in the financing culture toward short-term variable-interest loans, instead of long-term fixed-rate loans, motivated by banking supervisors will transfer the management of interest rate risks to the credit clients. The effects of such a supervision policy would be different from country to country. The German corporate finance market is primarily focused on the bank loan booked in the bank balance sheet but the Anglo-Saxon companies, on the other hand, are primarily capital market oriented. In addition, floating-rate loans are far more important in the U.S. or in the U.K. than in Germany. From a cost perspective, it is not advisable for most German SMEs to place a corporate bond on the money and capital market instead of asking the bank for a loan; since only when raising millions of euros in the upper double digit range does raising capital in the capital markets become economically viable for SMEs [Hausschild and Kral (2013)].

The risk of interest rate changes is not the only factor relevant for the assessment of financing in an economy. The NSFR also has a negative impact on long-term financing for fixed-term loans. Credit cooperatives, such as the Volksbanken and Raiffeisenbanken, will have to pass on the costs of intensified regulations in case of interest rate risks to the customers.

The creation of a common “level playing field” with international standards in regulation is, on the one hand, to be welcomed. However, the simplification of rules can quickly lead to a pragmatic but unrealistic “one-size-fits-all” approach. There is no doubt that the competitiveness of SMEs varies widely from country to country. Financing cultures do differ historically and borrowers vary dramatically in figures like average size, internationality, equity ratio, growth or RoE. An undifferentiated harmonization of the regulatory system can lead to the destruction of long-term financing structures in Germany. Capital adequacy for interest rate risks and the introduction of the NSFR would certainly change the financing habits of SMEs in Continental Europe.

Table 1: Annual proportion of interest income resulting from term transformation

BANKING GROUP	2005	2006	2007	2008	2009	2005-2009
Private commercial banks	11.2%	6.2%	1.8%	1.4%	8.7%	4.6%
Savings banks	25.8%	18.2%	6.5%	4.8%	24.9%	14.6%
Cooperative banks	23.5%	16.8%	5.9%	4.7%	24.8%	12.7%
Other banks	21.3%	15.4%	5.6%	2.9%	13.5%	8.7%
All banks	23.8%	16.9%	5.9%	4.6%	24.3%	12.3%

Memmel (2010)

7. CONCLUSION

The "one-size-fits-all" approach to regulation places small credit institutions at a disadvantage compared to large credit institutions. Regulatory rules should take account of the differences in size between the individual banks, the focus of the business models, as well as country-specific characteristics in order to avoid distortions of competition. The integration of interest rate risks in pillar 1 of Basel III would have significant implications for credit cooperatives, for example, in Germany. Depending on the regulatory escalation stage regarding capital maturities, up to 86.9% of the credit cooperatives could have a minimum capital ratio of less than 10.5%. Consequently, capital requirements for interest rate risk in the banking book could change the long-term financing habit of fixed interest rates. The capital requirements within the framework of Basel III will worsen the banks' RoE, until its fully implemented, between 14 and 111 basis points per year, unless countermeasures are initiated.

References

Admati, A., P. DeMarzo, M. Hellwig, and P. Pfleiderer, 2013, "Fallacies, irrelevant facts, and myths in the discussion of capital regulation: why bank equity is not socially expensive," Graduate School of Business working paper no. 2065, Stanford University, http://stanford.io/2v0vs4T

BaFin, 2011, "Rundschreiben 11/2011 (BA)," Bundesanstalt für Finanzdienstleistungsaufsicht, http://bit.ly/2a8OkIT

BaFin, 2014, "Neujahrspresseempfang der BaFin 2014 - Rede von Dr. Elke König, Präsidentin der BaFin, am 16. Januar 2014 in Frankfurt am Main," http://bit.ly/2eYx4H5

Bank of England, 2014, "Financial Stability Report," issue no. 36, http://bit.ly/2uHLpl9

BCG, 2011, "Risk report 2011 – facing new realities in global banking," Boston Consulting Group, http://bit.ly/2v0LHi0

Becker, H. P., and A. Peppmeier, 2011, Bankbetriebslehre (8. Aufl.), Herne: Kiehl

Berg, T., and M. Uzik, 2011, "Auswirkungsstudie Basel III: Studie im Auftrag des Bundesverband mittelständische Wirtschaft," http://bit.ly/2w9DRb6

BIS, 2006, "Internationale Konvergenz der Eigenkapitalmessung und Eigenkapitalanforderungen. Überarbeitete Rahmenvereinbarung," Bank for International Settlements, http://bit.ly/2uldhW5

BIS, 2010a, "An assessment of the long-term economic impact of stronger capital and liquidity requirements," Bank for International Settlements, http://bit.ly/2vD3Whr

BIS, 2010b, "Assessing the macroeconomic impact of the transition to stronger capital and liquidity requirements," Macroeconomic Assessment Group, Bank for International Settlements, http://bit.ly/2vD3Whr

Bonner, C., 2012, "Liquidity regulation, funding costs and corporate lending," DNB working paper nr. 361, http://bit.ly/2uOnU5F

Bridges, J., D. Gregory, M. Nielsen, S. Pezzini, A. Radia, and M. Spaltro, 2014, "The impact of capital requirements on bank lending," Bank of England working paper nr. 486, http://bit.ly/2wPwQcA

CEBS, 2010, "Results of the comprehensive quantitative impact study," Committee of European Banking Supervisors, http://bit.ly/2v0z5Yz

De Nicolò, G., A. Gamba, and M. Lucchetta, 2012, "Capital regulation, liquidity requirements and taxation in a dynamic model of banking," IMF working paper nr. 72, http://bit.ly/2uOk6l3

Deutsche Bundesbank, 2006, "Bestimmungsgründe der Zinsstruktur - Ansätze zur Kombination arbitragefreier Modelle und monetärer Makroökonomik," Monatsbericht April, http://bit.ly/2i5uinv

Deutsche Bundesbank, 2011, "Basel III – Leitfaden zu den neuen Eigenkapital- und Liquiditätsregeln für Banken," http://bit.ly/2v0ZQvT

Deutsche Bundesbank, 2012, "Die Rolle des „Baseler Zinsschocks" bei der bankaufsichtlichen Beurteilung von Zinsänderungsrisiken im Anlagebuch," Monatsbericht Juni, http://bit.ly/2vHChtL

Deutsche Bundesbank, 2013, " Bankinterne Methoden zur Ermittlung und Sicherstellung der Risikotragfähigkeit und ihre bankaufsichtliche Bedeutung," Monatsbericht März, http://bit.ly/2wBrZ0t

Deutsche Bundesbank, 2014, "Finanzstabilitätsbericht 2014," http://bit.ly/2w00iiE

Deutsche Bundesbank, 2015, "Ergebnisse des Basel III Monitoring für deutsche Institute," http://bit.ly/2i4OEgl

Dietrich, A., G. Wanzenried, and K. Hess, 2013, "The good and bad news about the new liquidity rules of Basel III in Western European countries," Journal of Banking and Finance 44, 13-25

Elliott, D. J., 2010, "Quantifying the effects on lending of increased capital requirements," http://brook.gs/2v111eN

EBA, 2014, "Results of 2014 EU-wide stress test," European Banking Authority, http://bit.ly/1tzlNyi

Fischer, M., and D. Heil, 2015a, "Negative Folgen für die Kreditvergabe," Profil – das bayerische Genossenschaftsblatt, 2(2015), 26–28

Fischer, M., and D. Heil, 2015b, "Zinsänderungsrisiken aus dem Anlagebuch und Eigenkapitalunterlegung für Banken," http://bit.ly/2w0y3Qx

Fischer, M., and S. Lanz, 2004, "Finanzkennzahlenanalyse bei Banken – zwischen Erkenntnis und Illusion," in Fischer, M., (Hrsg.), Wertmanagement in Banken und Versicherungen (S. 355-391). Wiesbaden: Gabler

Hausschild, S., and S. Kral, 2013, "Mittelstand setzt weiter auf Bankkredit," Betriebswirtschaftliche Blätter

Kashyap, A. K., J. C. Stein, and S. Hanson, 2010, "An analysis of the impact of "substantially heightened" capital requirements on large financial institutions," http://hbs.me/2i4VS43

King, M. R., 2010, "Mapping capital and liquidity requirements to bank lending spreads," BIS Working Paper nr. 324, http://bit.ly/2vHJytA

Marchesi, M., M. P. Giudici, J. Cariboni, S. Zedda, and F. Campolongo, 2012, "Macroeconomic cost-benefit analysis of Basel III minimum capital requirements and of introducing Deposit Guarantee Schemes and Resolution Funds," JRC Scientific and Policy Reports, http://bit.ly/2v0NKTj

McKinsey 2010, "Basel III and European banking: its impact, how banks might respond, and the challenges of implementation," McKinsey working paper on risk nr. 26, http://bit.ly/2i5yV0M

Memmel, C., 2010, "Banks' exposure to interest rate risk, their earnings from term transformation and the dynamics of the term structure," Deutsche Bundesbank discussion paper nr. 7, http://bit.ly/2uO6rup

Miles, D., J. Yang, and G. Marcheggiano, 2011, "Optimal bank capital," Bank of England discussion paper nr. 31, http://bit.ly/2vCLO6Z

Österreichische Nationalbank, 2008, "Leitfaden zum Management des Zinsrisikos im Bankbuch," http://bit.ly/2w120VT

Reifner, U., D. Neuberger, R. Rissi, and S. Clerc-Renaud, 2011, "CRD IV – impact assessment of the different measures within the Capital Requirements Directive IV," Directorate General For Internal Policies Study, http://bit.ly/2x3JYdn

Schätzle, D., 2014a, "Impacts of Basel III capital regulation to German cooperative banks - an empirical analysis based on a balance sheet simulation," http://bit.ly/2w134El

Schätzle, D., 2014b, "Die Auswirkungen von Basel III auf Genossenschaftsbanken. Eine Betrachtung der Kernkapitalquoten und Liquiditätskennziffern sowie eine empirische Analyse auf Basis von Bilanzsimulationen," in Theurl, T. (Hrsg.), Münstersche Schriften zur Kooperation, Band 110. Aachen: Shaker

Slovik, P., and B. Cournède, 2011, "Macroeconomic impact of Basel III," OECD Economics Department working paper nr. 844, http://bit.ly/2vDsZks

Schierenbeck, H., M. Lister, and S. Kirmße, 2014, Ertragsorientiertes Bankmanagement (9. Aufl.). Wiesbaden: Springer Gabler

Schmale, I. (2012). Genossenschaften und soziale Marktwirtschaft in einer pluralistischen Gesellschaft. Wege zu einer gerechten und leistungsfähigen Entwicklung, in Ringle, G., and H. Münkner (Hrsg.), Genossenschaftliche Kooperation - anders wirtschaften! (Bd. 111, S.19-35). Marburger Schriften zum Genossenschaftswesen, Baden-Baden: Nomos

Schmaltz, C., S. Pokutta, T. Heidorn, and S. Andrae, 2013, "How to make regulators and shareholders happy under Basel III," Journal of Banking and Finance 46, 311–325

Voigt, K. I., and M. Fischer, 2016, Genossenschaftsbanken im Umbruch, Einfluss der Finanzmarktregulierung auf das Geschäftsmodell der Kreditgenossenschaften, De Gruyter Oldenbourg, Berlin.

Weigel, W., and L. Sierleja, 2015, "Besonderheiten bei der Rechnungslegung von Kreditinstituten und deren Auswirkungen auf die Bilanzanalyse und die Analyse von Kennzahlen," in Everling, O., and K. H. Goedeckemeyer (Hrsg.), Bankenrating (2. Aufl., S. 205-248). Wiesbaden: Springer Gabler

Table A1: Studies analyzed on the implications of regulations on financial institutions

#	AUTHOR	TITLE
1	King (2010)	Mapping capital and liquidity requirements to bank lending spreads
2	Schätzle (2014a)	Impacts of Basel III capital regulation to German cooperative banks – An empirical analysis based on a balance sheet simulation
3	Elliott (2010)	Quantifying the Effects on Lending of Increased Capital Requirements
4	Kashyap et al. (2010)	An analysis of the impact of "substantially heightened" capital requirements on large financial institutions
5	Bonner (2012)	Liquidity Regulation, Funding Costs and Corporate Lending
6	EBA (2014)	Results of 2014 E.U.-wide stress test
7	Schmaltz et al. (2013)	How to make regulators and shareholders happy under Basel III
8	BIS (2010b)	Assessing the macroeconomic impact of the transition to stronger capital and liquidity requirements
9	De Nicolò et al. (2012)	Capital Regulation, Liquidity Requirements and Taxation in a Dynamic Model of Banking
10	Brides et al. (2014)	The impact of capital requirements on bank lending
11	Dietrich et al. (2013)	The good and bad news about the new liquidity
12	McKinsey (2013)	Basel III and European banking: Its impact, how banks might respond, and the challenges of implementation
13	CEBS (2010)	Results of the comprehensive quantitative impact study
14	Deutsche Bundesbank (2015)	Ergebnisse des Basel III Monitoring für deutsche Institute
15	Admati et al. (2013)	Fallacies, Irrelevant Facts, and Myths in the Discussion of Capital Regulation: Why Bank Equity is Not Socially Expensive
16	BCG (2011)	Risk Report 2011
17	Bank of England (2014)	Financial Stability Report
18	Berg and Uzik (2011)	Auswirkungsstudie Basel III – Die Folgen für den deutschen Mittelstand
19	Marchesi et al. (2012)	Macroeconomic cost-benefit analysis of Basel III minimum capital requirements and of introducing Deposit Guarantee Schemes and Resolution Funds
20	Miles et al. (2011)	Optimal bank capital
21	Basler Ausschuss für Bankenaufsicht (2010)	An assessment of the long-term economic impact of stronger capital and liquidity requirements
22	Slovik and Cournède (2011)	Macroeconomic Impact of Basel III
23	Reifner et al. (2011)	CRD IV – Impact Assessment of the Different Measures within the Capital Requirements Directive IV

Source: Voigt and Fischer (2016)

Transforming the theory and practice of risk management in financial enterprises

TOM BUTLER | Professor, GRC Technology Centre, University College Cork, Ireland
ROBERT BROOKS | Director, Risk Advisory, Deloitte, London, UK

ABSTRACT

This paper highlights the problems facing financial institutions in managing risk at an enterprise level. Chief risk officers (CROs) are confronted with the significant task of managing risk due to the high degree of uncertainty over the provenance and accuracy of risk data and information. This paper, therefore, considers the following questions:

- What is required to provide the group risk function with the same level of oversight and control over risk data and information that enterprise resource planning (ERP) systems have provided group finance?
- What is required for the wholesale transformation of risk management in the enterprise?
- How do business operating models need to change to facilitate true integration of business objectives and related risks?

While the problems with the siloed nature of risk management have been noted, the final point above is concerned with the disconnection between the management of business objectives and that of risk. The fundamental question that this article aims to answer is: How can GRC (governance, risk management, and compliance) practice and systems evolve to support the integration of risk management with business management?

1. INTRODUCTION

The banking industry is, in our opinion, at a cross-roads in terms of how banks address the challenge of navigating between the **Scylla** and **Charybdis** of regulatory compliance and enterprise risk in order to maximize shareholder wealth, while also meeting the expectations and information needs of an increasingly diverse set of stakeholders. We can see from the deluge of fines and other penalties levied by regulators in recent times [CCP (2015)] that some banks appear to be following **Odysseus** in choosing between what they consider to be the lesser of two evils – that is, avoiding grappling with the swirling whirlpool of enterprise risk while navigating the Messinian financial straits.

Yet, others appear to be oscillating between the rock and a hard place in terms of meeting the challenges of regulatory compliance and addressing the complex, paradoxical issue of enterprise risk, without doing either to the satisfaction of regulators or stakeholders. While banks appear to be willing to incur regulatory fines, accepting the recent volatility in global banking stocks is something else altogether, as shareholder wealth is being steadily eroded. Some now argue that the problem of addressing enterprise-wide risk effectively, and with due reference to regulatory requirements, can guarantee safe passage through these dire straits. In navigating this course, it is the visible hand of the CRO that needs to be on the tiller. Thus, it is in the hands of the CRO, as the bank's **First Officer**, that the safety of the financial institution lies in today's uncertain environment [Mikes (2008)]. Indeed, the same could be said of the banking industry, where systemic risk is concerned.

This paper considers the challenges the CRO faces in managing organizational risk in a highly-regulated industry. The management of enterprise risk is a complex activity, and a CRO may be forgiven for envying his fellow C-suite colleagues, whose tasks are not as onerous in informational terms, or equivocal in terms of internal and external expectations. It is significant that while there is "an abundance of principles, guidelines, and standards" and "risk management is a mature discipline with proven unambiguous concepts and tools," Mikes and Kaplan (2015) argue "that risk management approaches are largely unproven and still emerging." This applies, in particular, to the management of enterprise risk. Hence, the challenges facing the CRO are considerable. However, the CRO's role is complicated considerably by the paradox that banks are inherently risk-takers – as risk-taking is an essential part of business activity in financial institutions, more so than any other. In the absence of an enthusiasm for taking risks, the types of rewards valued by stakeholders (including internal actors, such as traders) and shareholders, in particular, will not accrue.

"The management of enterprise risk is a complex activity, and a CRO may be forgiven for envying his fellow C-suite colleagues, whose tasks are not as onerous in informational terms, or equivocal in terms of internal and external expectations."

A riskless bank is a logical contradiction, a dysfunctional institution that will be as doomed to fail as its opposite. The CRO in the riskless bank sees all risks as bad. This is problematic as the baby of good risks is often thrown out with the bathwater of bad risks [Stulz (2015)]. In considering the nascent body of research on such matters, we argue that the role of the CRO is to work with C-suite colleagues to maximize the opportunity for good risk-taking, with profitable outcomes, while minimizing bad risk-taking and associated losses, including regulatory penalties.

2. WHY WE NEED TO RETHINK HOW RISK IS MANAGED IN THE ENTERPRISE

In the 1990s, the finance function in business enterprises underwent a transformation through the adoption of ERP systems, which later became enterprise systems. The business driver for this transformational change was the need to gain control over the finances of large corporations by removing the duplication of effort in financial accounting across business units. Financial and accounting data was, like today's risk data, stored in data silos dispersed across the enterprise. This made the production of enterprise-level financial statements problematic, it also made internal and external auditing extremely difficult. ERP-enabled reengineering and transformation of financial audits considerably enhanced transparency and control of enterprise-wide financial and cost accounting to the chief financial officer (CFO) [Morris (2011), Chang et al. (2014)].

The need to automate the auditing of enterprise risk has driven the management of processes to control

risks within banks. Consider, for example, that controls testing is typically being employed to manage the various categories of operational risk, including IT risk, business resilience, and so on. This has clear efficiency gains for banks that automate and align control processes globally. In this regard, GRC tools are being employed to transform risk management functions and they continue to be invaluable for this purpose. However, there is a realization that financial enterprises need to transcend the process automation perspective and look at the problem of risk management in a different light.

Several questions present themselves for consideration at this point. Why, for example, would one wish to consider risk in an integrated way? What benefit does an integration approach offer, when it is widely accepted that risk management is best carried out by the first line of defense? What are the implications for risk management when, as Argyris (1976) argues, "espoused theory" in an organization is commonly at odds with the "theory in use"?

One "espoused theory" in common currency is that it is the first line of defense, usually operational management, that owns and manages the risks in an enterprise. Consequently, operational managers are accountable for applying corrective actions to address deficiencies in processes and controls [Sadgrove (2016)]. In other words, operational managers are, or are expected to be, responsible for identifying and assessing risks, as well as devising, applying, and supervising effective internal controls, while also ensuring that risk and control procedures are operationalized. In summary, the chief "espoused theory" in business enterprises is that operational managers should identify, assess, control, and mitigate risks in a manner that is consistent with their goals and objectives and those of their organization.

The problem here is that this can only be achieved if the commitments of such managers are aligned with corporate and regulatory objectives. However, when it comes to the first line of defense, "espoused theory" is typically at odds with the "theory in use" [Evans and Quigley (1995)], as the recent Wells Fargo fiasco on cross-selling indicated [Back (2016)]. In March 2017, Toronto Dominion Bank lost over CAD7 billion of its value as news reports revealed how bank employees were under pressure to sell inappropriate products to customers. Interestingly, in a statement that is indicative of a defense of "espouse theory," the bank disputed the reports and stated that "the environment described in the media report is very much at odds with how we run our business, and we don't recognize it from our own perspective, experience or assessments" [CBC News (2017)]. This paper offers theoretical and practical insights into how such problems can be effectively addressed.

2.1 Controllable risks

In delineating our thesis, we first focus on controllable risks that are non-financial in character. In our conceptual schema, a controllable risk has the following attributes:

- It is relevant to the achievement of a business objective.
- It is knowable.
- It is survivable.
- It is capable of being influenced by management action.

Thus, we argue that risk events should be able to bring about a desirable outcome or business objective, otherwise how are they distinguishable from random events? The problem here occurs when managers attempt to consider all possible events that could lead to the non-achievement of a valid business goal or objective, which maximizes shareholder wealth and is compliant with regulatory requirements. Since managers, and in particular senior executives, always operate under incomplete knowledge, their rationality is bounded [Simon (1955)]. Consequently, managers typically "satisfice" and adopt a general risk mitigation strategy of "holding capital" [Altunbas et al. (2007)].

Uncertainty and incomplete knowledge is the reason why risk events are often unrecognized or ignored [Taleb (2005)]. However, it may simply be that managers are not able to identify such events as risks, in which case they are overlooked. Alternatively, if managers increase their knowledge of risks and improve their detection capabilities, risk events can lose their ability to influence business outcomes over time. Of course, risk events must be survivable, if individuals and organizations are to learn from them and prepare for the next occurrence. If risk events are identified but uninfluenceable by management action, then managers either accept the risk or remove the related business objective.

It is logical to conclude that in order to control a risk one must first understand it. Hence, the first and second line of defense in a financial institution need to acquire, manage, and apply knowledge about the

business, its objectives, its environment, and the risk itself. While a business objective can be readily identified and known, information about the risks that threaten the achievement of a business objective, and the risks that the business faces once the objective is finally achieved, is not always readily available. This is, therefore, the principal challenge facing business managers in financial institutions. The following section helps address this problem.

3. RETHINKING HOW WE CONCEPTUALIZE RISK

When it comes to certain categories and sub-categories of risk, there is an important business imperative to manage them, as they tend to be predominantly in the bad risk category. We are not referring to risks that may be good at an individual or a unit level, but bad for the enterprise, as they may collectively exceed its appetite for risk-taking. Examples of such risks are business transactions undertaken by traders to maximize their own returns, but that, as a consequence, place the enterprise at risk. Such matters are equivocal in terms of their acceptance by business, and need to be addressed on an individual basis by managers, or prohibited by business rules. Examples of unqualified bad risks, which may be associated with the principal-agent problem, generally fall into the operational or conduct risk categories [Alexander (2006), Jarrow (2008)].

To be able to manage risk better at the enterprise level we need to reexamine risk in a fundamental way. A central plank of our thesis is that a model of risk, and its categorization, is required that reflects the human and organizational realities of risk management in the enterprise. This is particularly true where operational and conduct risks are concerned in financial enterprises.

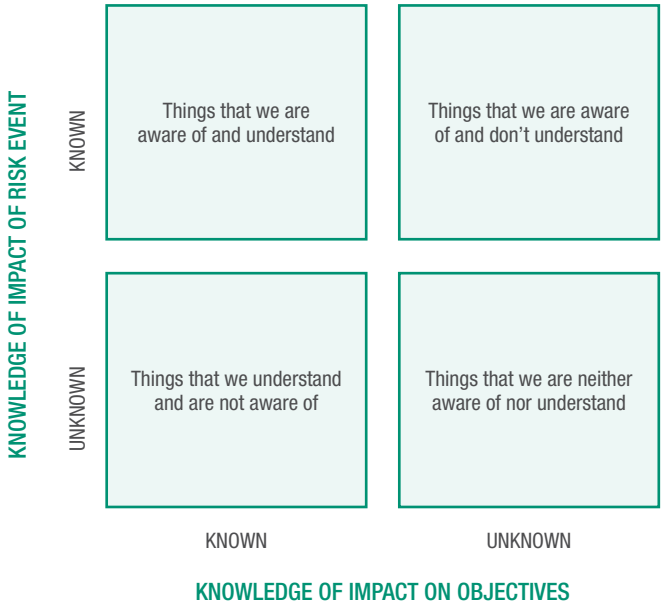
Using the ISO/IEC (2002) guide's definition of risk as our starting point, we conceptualize risk as the "effect of uncertainty on objectives," with the important elements of this definition being "effect," "objectives," "uncertainty," and the 'event' to which we are referring.

- An "event" is the cause that has an "effect."
- "Objectives" are the things that are being impacted.
- "Uncertainty" is the level of (or absence of) prior knowledge – it is not, in this case, an estimate of probability.

"The future is uncertain." While research in quantum mechanics disproved the existence of a deterministic world, some certainties or near certainties about the future in business can, in fact, be deduced. Unplanned events do occur, bringing both good and bad risks with them, and the mere act of planning can uncover the existence of at least some of them. However, successful planning depends, in large part, on the prior knowledge, experience, and expertise of the planner and the complexity ("predictability") of the internal and external environment.

To make our reconceptualization of risk more tractable, we adopt the Johari Window from the field of cognitive psychology [Luft and Ingham (1961)]. This has been used as a tool to enable self-awareness, knowledge, and understanding in several domains, such as in the defense [Petraeus (2015)] and national security sectors in the U.S. (and made famous by Donald J Rumsfeld in his response to the US DoD on Feb 12th 2002) or in thinking about operational risk [Kim (2014)]. Keeping with the Johari Window's 2 x 2 matrix presented in Figure 1, and considering "knowledge of impact of risk event" and "knowledge of impact on objectives" axes, the figure is instructive. Take, for example, the fact that in many organizations risks that can be managed and controlled are often considered to be unmanageable and uncontrollable. Operational and conduct risks, which should be considered known-knowns (KK), and fully managed and controlled, are typically ignored or categorized as either unknown-knowns (UK), known-unknowns (KU), or, worse still, unknown unknowns (UU),

Figure 1: Risk awareness model (RMA)



for a variety of reasons. The primary causes are related to a dearth of information or, rather, the existence of information asymmetries [Abraham and Cox (2007), Liao et al. (2009)].

In order to bring greater clarity to how risk is conceptualized, financial organizations should, in our opinion, conceptualize risks, risk factors, and risk events into four basic categories:

1. UU: these are the risk events that firms do not know about, and only identify them ex-post – these are referred to as "black swans."

2. KU: are risk events that firms know about, but cannot anticipate, understand fully, or quantify.

3. UK: here, the impact on enterprise objectives are known, but the specifics of risk events (because they may be novel in nature) are not known. In addition, because of the siloed nature of financial institutions, or the existence of information asymmetries, such risks cannot be comprehensively quantified at an enterprise level.

4. KK: are risk events that firms know about and where the impact on objectives are fully known, and, therefore, can be quantified, mitigated, or fully eliminated.

Some key conclusions can be drawn from this fundamental analysis:

- Risk cannot be effectively managed without first understanding related objectives. Without a clear understanding of objectives, all events are potentially risk events or, worse still, appear random until the impact is understood ex-post.
- Uncertainty can be reduced by knowing more about the events and their potential impact on objectives.
- There are two controllable elements of risk management that are not generally considered by organizations when managing risk: 1) the setting of objectives; and 2) the acquisition of knowledge about the effect of risk events on objectives, and the potential events that could impact them.
- If individuals are to be responsible for managing risk on behalf of the enterprise, they need to understand how their objectives contribute to the organizational goals.

The implications of this analysis for the CRO and the enterprise risk function are manifold:

- Without a clear link between risk management and the achievement of objectives, too many manageable or controllable risks events are being placed into the UU category and managed by holding capital. The risks in this class should, ideally, be entered under the enterprise "residual risk" category and their consequent impact on objectives entered as "not-known." However, some will be Black Swan events and have significant impact on business objectives.
- Organizational complexity reduces the ability to enforce the link between individual objectives and those of the enterprise, hence events may be known, but their impact on enterprise objectives not known (KU).
- Having a clear understanding of business objectives, and how they cascade down through the organization, means that lead indicators can be created. This means that potential deviations (UK) can be detected and corrected in advance of enterprise objectives being impacted.
- Reducing organizational complexity so that the links between objective, action, and outcome are known (i.e., through process modeling), and potential points of failure monitored for events that could have an impact on objectives, enables organizations to eliminate or mitigate certain risks, such as KK).

Of course, cognition of objectives and related risk events are not necessary and sufficient conditions for a solution to the problem, organizations must apply that knowledge to actively manage/eliminate risk in a routine way.

A review of fintech and Risktech offerings indicate that IT-enabled enterprise risk management solutions are now sufficiently mature, and related technologies available, to permit firms to move from managing risks they consider to be UK, KU, and UU, due to information-related problems, and bring some of them into the KK category.

Taking operational risk as an example, it is evident that because of the complexity and uncertainty in identifying and quantifying risks associated with failed people, processes, and technologies, only a subset of operational risks are being effectively and efficiently managed as KK. This is happening even though it is an endogenous risk category, and the data already is, or can be made easily, available to manage it. With regards to conduct risk, it is evident that many aspects of wholesale, retail, personnel, and third-party conduct risks are manageable as data on risk events, losses, and related factors are available.

Given the recent pronouncements of the Basel Committee on operational risk and the World Economic Forum on conduct risk, firms will have to focus more on managing these two major sources of risk. WEF (2016), for example, states that conduct risk is "likely the largest single source of technologically-driven risk." BIS (2016) advocates a withdrawal of internal modeling for operational risk measurement capital and its replacement with a simplified standardized model. The implication here is that banks will have to adopt more granular and accurate approaches for identifying, classifying, mitigating, and controlling operational risk, if they are to come out on the right side of the proposed "standardized measurement approach." The only confounding issue relates to the presence of qualitative or unstructured data, much of which is the product of subjective human opinion that is open to bias, as indicated below. It only requires that readily available risk management technologies are applied to capture this data and transform it into knowledge, thereby making conduct risks, for example, knowable and actionable. As with operational risk, this is an enterprise-level problem that requires an enterprise-level solution.

4. BASIC PROBLEMS WITH THE CLASSIFICATION OF RISK

There are two schools of thought regarding management of risk in business, with the first viewing risk as being defined independently of business objectives and the second viewing it explicitly in terms of the achievement of organizational objectives [Bromiley et al. (2015)]. When business objectives are expressed quantitatively, such as in financial terms, it is a relatively trivial task to understand the relationships between management objectives at the base of an organizational hierarchy to those at the top. This is because in a quantitative, or financially-based, hierarchy there is a mathematical or formulaic relationship between entities, be it additive, subtractive, multiplicative, or through the application of fixed rules or formulae. Consolidation of the outcomes of business objectives is relatively straightforward, provided the data is available.

In this schema, formalization of organizational structures and processes, and the application of financial or management accounting standards, provide a consistency of classification. For example, profit or

cost centers reflect areas of ownership and control, while business units act as containers of profit centers. In this scenario, if all the known risks in financial statements are controlled, then all that remains are unknown external risks and/or human risks – failed people. The financial audit process, therefore, focuses on the existence and effectiveness of controls and residual risk is the subject of human judgment.

All this stands in stark contrast to the problems posed by risks that cannot be expressed in quantitative terms. Such risks are neither easy to aggregate or disaggregate. This is partly due to the classifications given to such risks, typically operational risks, which give rise to fraud, IT risk, conduct risk, legal risk, and so on.

The current conceptions of operational risk grew around the emergence and practice of risk professions. Thus, labels are accorded to different risk categories and sub-categories in the same way as a biologist might classify different species using taxonomies [Gallagher et al. (2005), Moosa (2007)]. Populating a risk taxonomy by classifying risks is a subjective activity and requires judgment based on a common body of knowledge and understanding within a profession [Blunden and Thirlwell (2012)]. Objectiveness in species classification was not available until the advent of DNA mapping. Objective classification using DNA shows the path and branches of evolution so that species, genus, family, order, class, etc., are accurately classified.

The objective classification of risk in financial services could show how risks are related and permit the identification of the causal chains that give rise to major risks. It could also illustrate where the "gaps" in empirical observations exist; it could also be employed to arbitrate between different subjective judgments or viewpoints.

There have been numerous attempts to classify risks in risk taxonomies. Take, for example, the approach of classifying risks in a taxonomy that disaggregates losses. The problem with this approach is that it is only satisfactory when the business or managerial objective is not to make a monetary loss. The problem with the "loss events" construct is that it is wholly quantitative or financial in nature, even if the loss events are often not modeled as such. In this schema, both cause and effect are typically expressed in financial terms, even though risk events that are not financial in nature may be the trigger for the event. For example, it might be reported that a £100 mln loss in the P&L was "caused" by 100 different loss events of £1 mln. This is probably true from a financial risk perspective (e.g., market, credit risk, etc.), where the efforts to manage the risk can focus on the loss event itself, using hedging or diversification strategies. However, the fact that each loss event is caused by a real event is ignored. This raises the possibility that future real events will not be detected. Managers rely on the assumption that each of these risk events impact the market and that, in aggregate, the impact on the market cannot be known. Where the risk event and the loss event can be linked directly, then attempts to manipulate the causal chain are positively discouraged – particularly when this leads to market abuse or insider trading. The exception to this is the action taken by a central bank in areas of current market manipulation, bond purchasing, and so on.

Where non-financial risk is concerned, active attempts to achieve an objective outcome by preparing ahead of time to prevent deviations from the outcome is the optimal way of managing risk. The only other alternative is to let the risk materialize and remediate it after the event. However, understanding the causal chain is critical, as it will ensure that managers take steps to avoid deviation from the trajectory required to reach intermediate goals and ultimate objectives by preparing for and negating risk events. This, however, is a costly approach to risk management.

The often used and least costly approach is to map the critical path and to design-out potential deviations, or to identify and mitigate locally any detected deviation. However, to be effective, this approach requires detailed process modeling; it also requires a better understanding of the type of risks under consideration [Rosemann and Zur Muehlen (2005)]. There is little evidence that either conditions are being met in practice. We turn next to this topic, which builds on the RMA presented in the previous section.

4.1 Characteristics of knowable and controllable risks

In order to begin to address the above problems with a risk classification approach, we extend the conceptualization of our RMA by defining the characteristics of knowable and controllable risks. First, they are **additive**: examples are accounting risks related to debtors' ledger, creditors' ledger, etc. These are factors that can be measured objectively. Second, they are **auditable**: knowable and controllable risks rely on the "chain of custody" of information to manage them. This approach relies on the fact that there is an

immutable truth at the start of the chain that can be traced to an output, without manipulation on the way. If the entire chain of evidence is within an organization, its validity can be verified. Problems occur, however, when the chain crosses organizational boundaries. Third, **compound** risks are those that are insignificant in relative scale, at the bottom-tier of the organizational hierarchy, but become problematic when they interact with other categories of risks, and exert an enterprise-level effect. Fourth, **singular** risks that impact business objectives to the same extent, wherever they occur in the organization (e.g., reputational risks such as LIBOR manipulation). Such risks are characterized by a separation of the owner of the risk and the actor(s) from which the risk emanates – for example, the LIBOR manipulation resulted in the boards of firms having to take ownership, even though the "causal owner" was much lower in the hierarchy. Fifth, **poolable** risks, such as IT risk, which is a pooled risk as it requires particular levels of expertise across both IT and business functions. Here, managers need to possess specific levels of domain knowledge to understand such risks.

Another category of knowable risks is, in our opinion, neither controllable nor easily detectable. We know they can occur because they have occurred previously, but they are not predictable. Sub-categories here include **internal risks**, such as employee risks, emanating from poor judgment, criminal intent, reckless behaviors, negligence, incompetence, and so on. In addition, there are **external risks**, such as customer risks, where the chain of evidence for audit begins outside of the organization.

Then there are risks that are unknowable due to uncertainty. These usually have an impact on an organizations' **survival objectives**. Such "black swan" risks may lead to the physical cessation of business. Risks in this category include solvency-related risk events that occur when decisions taken inside or outside of the organization have a domino effect and impact on a firm's ability to trade. Such risk events may originate in, for example, a decision to delay payment to creditors, a breach of trust, or reputational damage with stakeholders, and so on. Responses to such risks depend on operational resilience, or reality antifragility, as Taleb (2012) puts it.

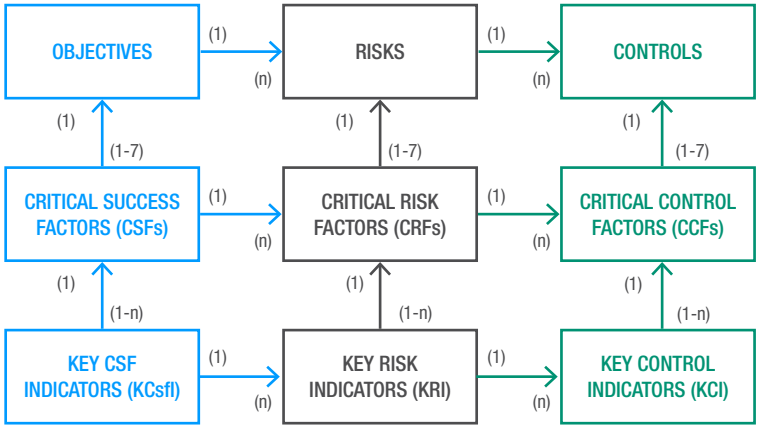
For all these reasons, we argue that financial organizations need to manage risk in the context of business objectives and transcend the tendency to silo risk while also separating and divorcing business and risk management processes. It is to this topic that we now turn.

5. AN INTEGRATIVE APPROACH TO MANAGING BUSINESS OBJECTIVES AND RISKS

Risks at all levels in an enterprise should be linked to the achievement of related organizational objectives. However, there is little guidance in the academic or practitioner literatures on how to achieve this. There are certainly a wealth of complex standards, frameworks, and methodologies that purport to help practitioners manage risk, however, in our opinion, these are either too narrow, too fragmented, or are simply too labyrinthine, resulting in practitioners becoming lost in the detail and failing to realize the benefits. Moreover, none provide the type of informational capabilities to serve as a model for the form of enterprise-wide risk management system required by CROs to serve the information needs of the C-suite or the boards of financial institutions.

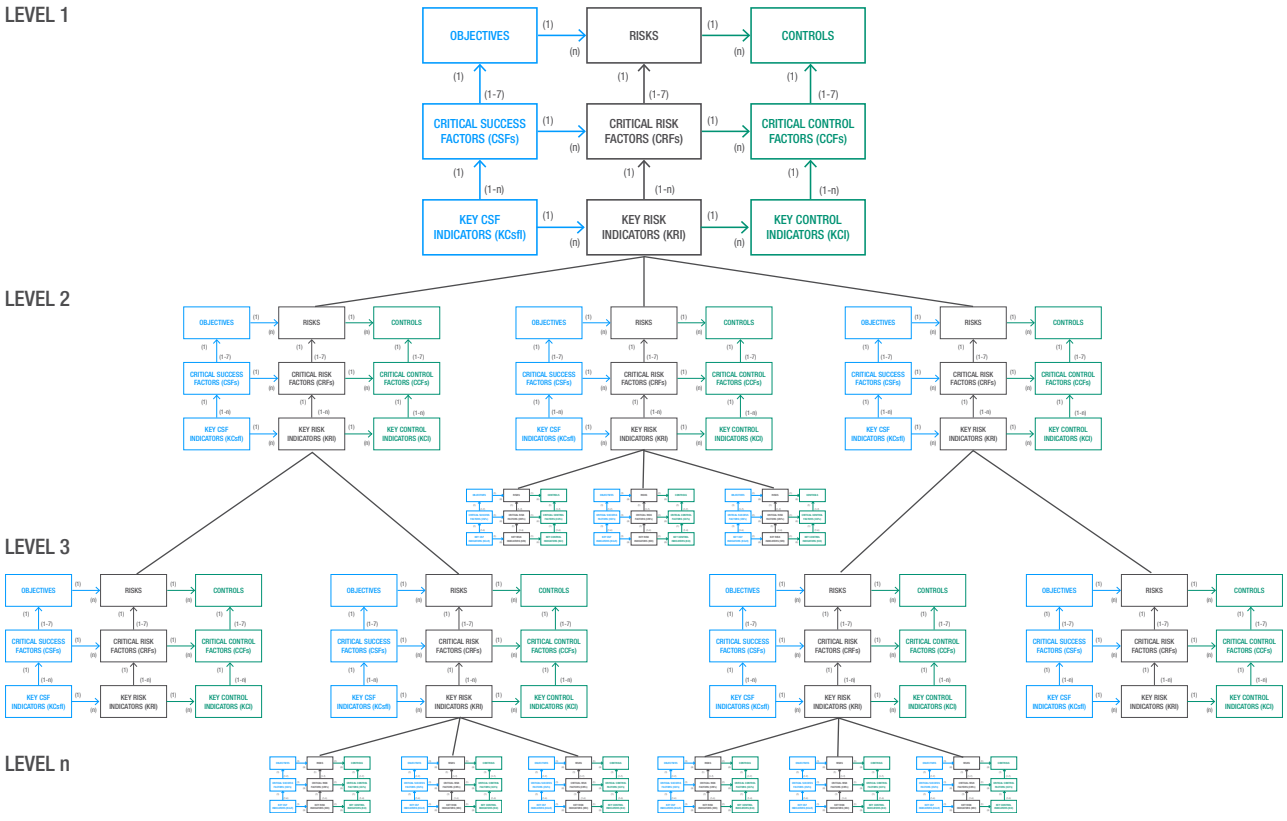
This paper draws on seminal work of Rockart (1979) on

Figure 2: A CSF-based model on linking business objectives, CSFs, risks, and controls



"critical success factors" (CSFs) and Kaplan and Norton (1996) on the "balanced scorecard" to present insights into how all this can be achieved. Figure 1 illustrates our perspective. Rockart (1979) defines critical success factors as "the limited number of areas in which results, if they are satisfactory, will ensure successful competitive performance for the organization. They are the few key areas where "things must go right" for the business to flourish. If the results in these areas are not adequate, the organization's efforts for the period will be less than desired."

Figure 3: Hierarchy of objectives, risks, and controls



CSFs are different from objectives and goals. Objectives are general statements about the directions in which a firm (sub-unit or manager) intends to go, without stating specific targets to be reached at particular points in time. Goals are the specific targets that are intended to be reached at a given point in time by managers. A goal is thus an operational transformation of one or more objectives. Hence, a manager's goals are the targets that they will aim for.

CSFs are the key areas of activity that most influence success or failure in their pursuit of goals and related objectives. A CSF is what has to be done in order to achieve a particular goal and a related objective. Goals and objectives are the ends, while CSFs are the means to those ends. Figure 2 illustrates this relationship graphically. However, it is clear from the literature that the CSFs-goals-objectives construct is rarely adhered to. For simplicity's sake, we conflate goals and objectives to simply objectives, as indicated in Figure 1. Following Rockart, our schema posits that each objective has 1-7 related CSFs. Likewise, each CSF has 1-n key indicators. Mapping this schema onto the risks and controls dimensions results in similar cardinalities. It is clear that each objective may have 1 or more (n) risks, while each risk may have 1-n controls. This type of relationship also exists at the critical factors and key indicators levels, as indicated in Figure 2.

In Rockart's schema, CSFs and objectives are influenced by "problems" – i.e., business problems to be solved. Attaining business objectives involve undertaking risks. From a risk management perspective, this indicates the existence of "critical risk factors" (CRF), the presence of which influence the attainment of CSFs. We also believe that just as CSFs may be decomposed into measures, such as "key CSF indicators" (KCsfI), CRFs may be decomposed into measures such as "key risk indicators." Usually, the intermediate modeling of CRFs is omitted with risks simply being mapped to KRIs. This omits an important analytical step, which could result in the omission of important risk indicators and poor measurement of risks. It is also clear that there may be relationships and overlaps between both sets of measures. Extending this model to include controls, we posit the existence of "critical control factors," which

are decomposed into "key control indicators" (KCIs). Unlike KPIs, KCsfIs, KRIs, and KCIs are lead indicators and, therefore, more relevant to the task at hand.

Figure 2, therefore, presents a normative, parsimonious model that captures the essence of how risks should be managed by business managers in the first line of defense. This approach of defining objectives, in terms of the CSFs that are required to meet them, in concert with the CRFs that impact the attainment of CSFs, and the controls required to mitigate risk events, appears to be a common-sense approach. This stands in contrast to the business as usual approach, where first line operational managers fail to identify, assess, control, and mitigate risks in a manner that is consistent with their objectives and those of their organization. Then, there are the risk and control frameworks and methodologies that purport to help practitioners, but are difficult to implement due to their complexity or ability to scale horizontally or vertically. Our research indicates that objectives and CSFs cascade in a hierarchy from top management down, while also spanning organizational units and functions.

Figure 3 illustrates the scalability of the proposed model, which incorporates qualitative and quantitative data. It is important that if this model is to be enabled by appropriate technologies, then it could provide both roll-up and drill-down capabilities, enabling risk data aggregation and enhanced risk management capabilities.

6. RISK MANAGEMENT AS DOUBLE LOOP LEARNING

Argyris (1976) argues that organizations typically engage in "single-loop learning," in that they generally apply fixed models for decision-making and problem-solving. In general, organizations rarely go beyond single-loop learning as a mode of behavior because they fail to question the assumptions underpinning their strategies or decision-making routines. Hence, they fail to develop what Aristotle calls practical or experiential knowledge. They also fail to build or evolve technical knowledge and skills. All this condemns organizations to apply the same decision-making and problem-solving routines over and over without learning how to improve their knowledge. This is a critical issue. Omerod (2007) illustrates that success in any area of endeavor is elusive because it is dependent on possessing appropriate knowledge about the organization, its business, and related risks. To achieve this, Argyris (1976) argues that organizations need to engage in "double-loop learning." This involves critical sense-making and subjecting governing assumptions and beliefs to question, the answers to which point to the need for new decision rules and the development of new routines. However, organizations also need to practice knowledge sharing, bridge knowledge gaps, and making learning outcomes explicit. We now consider this in the context of organizational knowledge and capabilities, and the management of business objectives.

The evolution of risk management with respect to business objectives depends on learning and knowledge acquisition. Knowledge of how to achieve an objective is related to the definition of CSFs for its attainment. However, we argue that there is also a need to identify CRFs in order to implement controls so as to ensure that the objectives-CSFs along the critical path are realized. It is important to understand objectives in terms of what is required to achieve them – i.e., CSFs and CRFs – otherwise actors are continuously confronted by decisions to consider all options over a range of paths in order to identify the next move.

Decision-making under uncertainty, caused by a failure to prepare, adds unnecessary complexity. The quality of the decision is a function of an agent's commitments, experiential knowledge, capabilities, intent, objectives, and a web of social and cultural conditions and factors [Nelson and Katzenstein (2014)]. However, as indicated above, managers typically "satisfice" on bounded knowledge and rationality [Kahneman (2003)]. If time and resources permit, managers may enact their decision theories through risk scenarios using focus groups, on one hand, or predictive modeling or simulation, on the other [Blunden and Thirlwell (2012)].

The research cited herein indicates that managers' decisions are typically based on experiential knowledge of critical success or risk factors expressed in the form of heuristics or routine decision patterns [Busenitz and Barney (1997)]. This offers a quicker route to the achievement of an objective. Typically, a manager defines a critical path, intermediate objectives, and then manages the deviations from these. Project management typically relies on such techniques. This requires effort and judgment to determine whether a deviation has occurred, and put a corrective action in place. Human judgement is augmented as the trajectory to the objective becomes known (estimated in reality). However, there is a need to focus on risk factors and

events that change the trajectory.

A **controllable risk approach** is possible through enhance learning and the development of new decision-making routines. In the proposed schema, all potential points of deviation from intermediate CSFs-objectives can be mapped and measured, and controls put in place to detect and correct deviations. Our model above also indicates a need to identify and measure CCFs as well as CRFs. This approach helps reduce the amount of human judgement required by adopting a "rule-based" approach to decision-making. Automated controls may then be employed for potential points of failure, as uncertainty and decision complexity is lowered.

7. ENTERPRISE INTEGRATION OF OBJECTIVE MANAGEMENT AND RISK MANAGEMENT

The optimal "risk appetite" models for operational risk tend to follow Kaplan and Norton's (1996) "balanced scorecard" approach. Here, the categories in the scorecard align to the objectives of the organization, they are (or should be, if the user is faithful to the model) linked to the CSFs for achieving them, and related measures are identified and operationalized. However, this is not the norm. In addition, organizations, whether they use the balanced scorecard or not, typically create a risk taxonomy that stands separate from, and is not integrated with, organizational strategies or objectives, if they are indeed defined and codified.

This paper argues that organizations should be developing and managing a taxonomy(ies) of business objectives that are integrated with risk and control taxonomies. In this scheme of things, the macro objectives are expressed in a taxonomic hierarchy, different levels and branches of which are owned by appropriate managers and units within the organization. Based on research and practice, we now briefly examine a general framework of business objectives using a basic taxonomy.

Enterprise risks are those events that impact the upper levels of the objectives hierarchy presented below. Addressing such risks are influenced by organizational knowledge and capabilities.

Level 1: Survival objectives

- License for business and trading.
- Ability to trade (sell).
- Ability to buy inputs (buy).
- Ability to staff.
- Ability to direct and coordinate resources.
- Resilient to external events.

Level 2: Strategic objectives

- Ability to satisfy stakeholders (valuation, profit motive).
- Financial integrity.
- Solvency.

Level 1 and 2 capture an organization's survival and strategic objectives. The subsequent levels in the objectives taxonomy should map to the organizational structure (extending the levels already described).

Objective and risk decision-making on more strategic and tactical objectives will clearly be the province of higher levels in the organizational hierarchy. And where knowledge and expertise to inform decision-making exists elsewhere in the organization, sufficient governance needs to be put in place so that the owner of the objective is responsible and accountable for decisions.

Assuming that an enterprise risk management system exists to manage objectives (according to the model in Figures 2 and 3, for example), what supporting roles would such a system be expected to perform? We argue that an enterprise risk management system should:

- Provide the integrative capabilities for managers at all levels to define their objectives, CSFs, and associated measures.
- Identify the related risk events and map these to CRFs and measures.
- Monitor and manage the taxonomy of objectives and map them to the correct risk nodes in the hierarchy.
- Define controls for risks, and CCFs with associated measures where appropriate.
- Confirm that controllable risks are controlled, and that the control environment is heathy.
- Provide assurance that:
  - Decision making is made at the right level of the organization.
  - Good decisions are being made.
  - Bad decisions being made as a result of incorrect information is eradicated.

- Minimize the existence of compound risks.
- Monitor the existence of singular risks and ensure that correct governance is in place for them.
- Provide internal audit capabilities that monitors the "healthiness" of the pooled risk.
- Provide a heatmap/dashboard to the C-suite and board indicating the levels of risk resilience.

8. ON THE RELATIVE INFANCY AND IMMATURITY OF ENTERPRISE INFORMATION SYSTEMS FOR RISK MANAGEMENT

It is clear that the chief risk executive of any bank is the CEO, not the CRO [Stulz (2015)]. The CEO of a financial institution with an appropriate risk culture will, however, value and leverage the particular knowledge and expertise of the CRO. Thus, research has indicated the importance of the relationship between the CRO and the CEO, and as the findings presented above illustrate, as far as enterprise risk is concerned CEOs need to listen to their CROs. Stanton (2012), who participated in the U.S. Financial Crisis Inquiry Commission (FCIC), found that the presence of "constructive dialogue" in a bank, and the inclusion of the CRO and risk perspectives in decision-making, was characteristic of those banks that avoided the type of losses that occurred in distressed or failed banks in the financial crisis.

Stanton (2012) argues that engaging in constructive dialogue will help empower CROs and the risk function to create the institutional framework – consisting of regulative, normative, and cultural dimensions and mechanisms – required to make the commitment to enterprise-wide management of risk a reality [Brandes et al. (2005)]. It is important to note that while the CRO and the risk team are responsible for identifying, quantifying, monitoring, and controlling risks, it is, as indicated, the function of the business to actually manage risks at operational, tactical, and strategic levels. Nevertheless, it is the responsibility of the CRO and the risk function to collaborate with the business to develop strategies, practices, and routines that are risk-optimal in terms of their profitability and contribution to shareholder wealth. Thus, the CRO has to balance the need to be objective and independent, with the requirement to collaborate with the business. To be a credible agent for change across the enterprise, it is vital for the risk function to be adequately informed, and it is here that information technology is, and will increasingly be, a vital source of hard data,

business intelligence, and management information. Unfortunately, the CRO and the risk function are not as well-served in this regard as other colleagues in the C-suite, such as the COO or the CFO.

8.1 Problems with information systems' support for the risk function

Financial, accounting, and transaction processing information systems in banks are highly mature in terms of their support for information and decision-making in the disciplines of finance and management. Such systems also help automate and enable reporting according to accepted standards, such as GAAP and IFRS. Thus, the CFO typically has at their fingertips the ability to determine the provenance of financial data and information through all levels and across functions in a bank – retail, commercial, or investment. The CRO is not as well-endowed, in terms of informational resources, as the CFO, as IT-enabled enterprise risk management systems are extremely immature, and comprehensive enterprise-level dashboard capabilities practically non-existent. Direct support for this contention comes from the Basel Committee on Banking Supervision's BCBS 239 principles for risk data aggregation [Grody and Hughes (2016)]. There are several reasons for this, which are worthy of reflection.

8.2 Problems of risk data completeness, accuracy, and quality

There is evidence that IT-enabled risk information systems are limited in a number of ways, particularly in terms of support for real-time risk measurement, monitoring, and control. Certainly, real-time risk measures exist for certain activities, but these tend to be silo-based. There is, unfortunately, a bigger problem. As risk management functions have evolved in banks, in particular lines of business and across the industry, areas of specialization have grown around the various categories of risk. This has led to fragmented risk management practices in terms of the application of approaches, capabilities, knowledge, procedures, and, of course, the manner in which risk data is managed and stored. Most significantly, the growth of banks and the digitalization of business has resulted in a proliferation of data silos. Thus, the data required to identify, monitor, and manage risk within and across business lines is stored in the databases of many hundreds of operational systems. The growth of this data is exponential, with new systems being introduced as banks digitalize their business [Tett (2010)].

Depending on the degree of autonomy in each business area within a bank's functional areas, risk executives typically employ unintegrated point solutions (often based on Excel spreadsheets) or risk management software applications developed by the IT functions in-house, or solutions from a range of vendors. However, the overall impact of often ad-hoc, unintegrated risk management systems at an enterprise level is for all intents and purposes negligible, due to their fragmented and siloed nature. In addition, there is an absence of agreed business vocabularies across many financial enterprises. Thus, business objects have multiple data representations, and data has multiple meanings attributed to them. Regulators find this situation extremely problematic.

> "Organizations should be developing and managing a taxonomy(ies) of business objectives that are integrated with risk and control taxonomies."

As a consequence of this, existing risk management systems also contribute to heightened operational risks, as business, IT, and risk professionals manually disambiguate, collate, and analyze business and related risk data. In situations where business lines have created data warehouses or data marts, and more recently data lakes, banks still find that the data is incomplete and unintegrated with key internal and external data. Consequently, accurate, consolidated measures of risk are rarely available for the entire bank or financial organization. Worse still, the provenance of data is problematic due to the manner in which data is governed by business and IT functions [Soares (2015)]. Thus, the CRO and business executives have problems in proving adequate data quality, lineage, and provenance to auditors and regulators, increasing regulatory risk and resulting in greater capital allocations.

8.3 Problems with risk models

The business assumption regarding the accuracy of existing tools and techniques for the identification and measurement of risk is, according to leading academics and practitioners, erroneous [Shojai and Feiger (2010)]. To illustrate this point let us examine the use of value-at-risk (VaR) as an enterprise tool for risk management. The first point to note, however, is that the data on

which VaR models are based must be of high quality, complete, and accurate, otherwise no matter how good the models are, they will produce inaccurate estimates.

VaR is used to measure a variety of risks, from an individual trading desk, to a business line, and on to a measure of corporate risk to be used by a CRO. However, there are significant limitations in using this approach. Building from a VaR for a particular unit or function, multiple VaRs may be combined to develop an enterprise-wide VaR for a bank. Correlations between the risks generated by different units may also calculated. Thus, it is possible, at least in theory, to estimate an overall measure of risk in a bank and to identify areas where risk appetite has been exceeded. In practice, however, there are problems in that VaR cannot be used to measure every risk and VaR models carry significant risks in themselves. Even when different categories and sub-categories of risks can be estimated using VaR, along with their correlations, a true measure of enterprise risk is not possible, as certain risks are not included and correlations estimated [Bamberger (2010)].

8.4 Fundamental behavioral and cultural issues

Then there are a range of more fundamental issues. We know from the work of Daniel Kahneman and others in the field of behavioral finance and psychology that economic actors operate under the influence of a raft of biases, which influence how they perceive risk. Such biases are difficult to identify and contaminate risk models generally assumed to be sound. Other biases and contaminants originate in the existence of competing commitments and moral hazard, where actors are incentivized to act in their own interest or short-term objectives, as opposed to that of their business unit or enterprise [Kegan and Lahey (2001)]. Then there is the nebulous matter of the culture of the bank or institution, which is extremely resistant to estimation or quantification, as are the mountains of qualitative or unstructured data collected and stored in a myriad of data repositories.

As indicated, the origins and emphasis of BCBS 239 reflects the current poor state of enterprise risk management across the industry, as risk data aggregation is, with few exceptions, wholly inadequate. Its principles provide a foundation on which the governance of risk in a banking enterprise can be based. Practitioners note that BCBS 239 does not require common risk metrics and challenge this notion by arguing that risk officers, business managers, and accountants need to architect finance and risk systems that are integrated and possess a common control and reporting framework. Be that as it may, the range of issues outlined above bear witness to the apparent intractability of the problems facing CROs in enabling the management of risk in and across the enterprise. The following section offers some direction in transitioning to next generation financial services.

9. NEXT GENERATION RISK MANAGEMENT IN BANKING

It could be concluded from the above that a CRO needs to possess similar strategic capabilities as those of a CEO, to understand regulations like a CLO, to know the business operations as well as a COO, to navigate financial risk similar to a CFO, to exhibit the same technical knowledge as a CIO, and understand risk data at the level of a CDO. Of course, if other C-level executives could view their business through the eyes of a CRO, then this would help simplify the organizational change that is required in the coming years.

We believe that financial institutions that do not recognize these basic realities will end up in deeper trouble than that which some of the major banks find themselves in at present. With traditional business models under pressure from new entrants and innovations from the fintech sector, and with margins squeezed from all sides, including regulatory compliance, banks will have no option but to take even greater risks. The CRO will make the difference here by enabling the business to identify and maximize the return on good risks and controlling, mitigating, or eliminating bad risks.

Given the regulatory forces and business drivers that currently shape their environment, financial institutions will need to rethink and transform not only their risk functions, but the status and role of the CRO. CEOs need to reorient their C-level teams to accept the risk function as a core business partner, and the CRO as business risk leader, if they are to transform and prepare their banks to face not only current challenges, but the all too certain future challenges and make their banks, as Nassim Taleb would say, “antifragile.” Information technology’s ability to transform organizations by automating their business process and informating

their people is a key enabler here [Zuboff (1991)].

9.1 INFORMATING AND AUTOMATING BUSINESS PROCESSES

Banks are no strangers to the transformational power of IT. IT-enabled software applications are being used to automate risky business processes, such as client on-boarding, KYC, and other customer-facing activities. Innovations in the fintech and regtech sector offer enhanced capabilities to informate and automate their activities across business lines. Digital innovations in e-banking/online/mobile banking, and so on, provide new avenues for automation and elimination of operational risks, such as failed people in anti-money laundering (AML). Utilities and regtech vendors offer a range of services to banks that can augment or replace inefficient and risky operations with tried and tested solutions, with, surprisingly, the support of regulators. Artificial intelligence [Castelli et al. (2016)], machine learning, blockchain [Jessel and Marshall (2016)], and robotics [Cocca (2016), Arwas and Soleil (2016)] are the new buzz words in an industry that is planning to automate, with virtual robots, certain middle and backoffice functions.

These are examples of the use of IT to minimize the need for manual processes across business activities and lines across the organization. Bad non-financial risks, such as operational risks, can be reduced or eliminated by simplifying, standardizing, and automating business processes, particularly where customers or partners in service delivery are concerned. Big data technologies are being used in concert with semantic technologies, predictive analytics, and machine learning to address a range of operational risks, from fraud, to insider threats, front running, and so on. Regtech-based semantic technologies are also being used to help legal, risk, and compliance teams deal with the mountain and complexity of regulations.

9.2 Navigating the digital labyrinth to manage and report on enterprise risk

As indicated, the core of many of the problems banks face in managing risk across the enterprise is the manner in which they manage data; both structured data isolated in siloed databases and spreadsheets, and unstructured data in documents and text fields. With few exceptions across the industry, this approach has seen little change since 2008. As indicated, BCBS 239 is heralding in a new era for risk data governance and risk data aggregation in banks large and small. The financial services industry generates more data and spends more on its storage than any other. Surprisingly,

there persists a basic inability to govern and manage that data, to interconnect it, link it with external information, and to make inferences from disparate and diverse data, wherever it exists. This makes risk management and compliance reporting hugely problematic and expensive. Manual data collation and integration remains the norm across the industry. This generally remains true for the global systemically important banks (G-SIBs).

The Enterprise Data Management Council (EDM Council) stated that the core problem was the absence of a common language or vocabulary within and across banks to describe the business meaning of data and metadata. The EDM Council is a global association of leading financial services organizations, technology vendors, and government agencies based in the US and Canada. The Council recognizes that a common language, enabled by semantic technologies, is required to better manage not only the mountains of data in and across banks, but also manage financial and systemic risks, and to enable comprehensive compliance reporting in the face of increased regulation. Thus, the EDM Council "co-opted" the software industry standards body, the Object Management Group (OMG), to collaborate in the development of, and to help institute, a standard vocabulary called the Financial Industry Business Ontology (FIBO) [Bennett (2013)]. While this is significant development at an industry level, individual banks need to develop related common languages to help add business meanings to, disambiguate, integrate, and link data internally and externally, be it structured or unstructured. Consequently, banks need to address what is the core problem for them and the industry: the absence of a common language to describe both business objects and processes and the risks attached to them. Since these are increasingly digitized, this means developing a common language for their data; one that bridges both business and IT functions of this data. There is also a need to arrive at agreed conceptions of the risks they face, that would, in turn, enable data integration and make risk data aggregation a reality. Thus, there is a need for a related common language for risk, expressed as risk taxonomies that are semantically enriched.

With few exceptions, the current fragmented offerings from the fintech sector are merely adding to the digital labyrinth, as new structured and unstructured data silos are being created. The same can be said of the budding regtech and risktech sectors in terms of offering comprehensive solutions for the particular problems faced by the financial services industry. In solving one problem, eliminating risk through process automation, others may be created.

The solution to the problem of what is a digital labyrinth is technically feasible and practically possible, given the rise of NoSQL technologies [McCreary and Kelly (2013)]. Unfortunately, there are few players in the market providing comprehensive solutions for the industry. One approach that is receiving much attention is data virtualization. This approach provides access to data directly from one or more disparate data sources, without physically moving the data, and presenting it in a form that makes the technical complexity transparent to the end-user. There is broad agreement across industry sectors that semantic metadata (based on the aforementioned common language) is required to make data virtualization and other NoSQL approaches work. Thus, semantically-enabled data virtualization will help underpin both enterprise risk management and enterprise risk reporting.

10. DISCUSSION AND CONCLUSIONS

In reflecting on the challenges facing CROs, we must return to the past to solve today's problems. As indicated, CSFs are those few things that must go well for an individual or an organization to ensure success in a business undertaking. We believe that the CSF method offers a tried and tested approach to rethinking how risk is managed at an enterprise level.

CROs and their risk teams would benefit in applying this tried and tested approach to identify their objectives, CRFs, and related data needs and information requirements. This seems sensible as complexity and uncertainty is the norm and the chances of developing an enterprise-wide risk dashboard remote if fundamental information needs are not formally defined and recorded. It would, for example, help CROs and their teams communicate their information needs to CIO/CTO/CDOs and the business. This is particularly important as information technology, be it fintech, regtech, or risktech, is being harnessed in an ad-hoc manner, with disintermediation of information by multiple systems adding to complexity and opacity of risk data in the CRO's office.

However, we need to go beyond current siloed approaches and apply the same methodology across the enterprise to help executive and managers at all levels, particularly those in the first line of defense, to create an

organizational taxonomy of business objectives, goals, CSFs, CRFs, and CCFs related measures. This should then be mapped to the standard risk taxonomies.

10.1 Reconsidering risk

Effectively managing risk still means we have risk. So, what is risk? What are the characteristics of risks, and why do we care about them? To recap, one cannot have risk without first having an objective to pursue. A risk is an event that may occur to prevent a business manager from achieving a particular objective. An objective could be something as general as being accepted by colleagues, or as specific as making profit on a derivatives deal. At an organizational level, objectives can either be considered as the aggregate of all of the objectives of the employees of the company, or employee objectives being a sub-categories of the objectives articulated by the executive committee.

In the finance function, the fact that objectives can be expressed in numerical terms means that the aggregation of financial objectives is achievable; objectives can be cascaded from top to bottom and activities, actions, and outcomes can be collated and aggregated in the same manner. Even large organizations can ensure that financial objectives are harmonized by using tools such as Finance ERP or modules in Enterprise Systems. As we stated above, problems occur when objectives cannot be expressed in numbers, or when nonfinancial conditions are imposed on those numbers – e.g., rules such as: "must not be from the proceeds of crime," "must not be from money launderers," and so on. The collective term for this type of risk is nonfinancial risk."

"Non-financial risk" (NFR) covers topics as diverse of reputational risk, cyber risk, compliance risk, operational risk, conduct risk, and legal risk. Each risk event may give rise to a loss event, but the risk itself does not represent a financial loss, unlike a market or credit risk. What is true of all NFRs is that if one prevents the risk event, the loss event is also prevented. What is also true of some NFR events is that if one can detect the risk event, one may avoid the loss. Which leads to the (not so) startling realization that the more one knows, the more time one has to prepare, and the more effective one is at preparing, the more likely it is that one will achieve the desired objectives. Thus, the significance of the points made previously for the need to develop double-loop learning.

10.2 How does GRC practice need to evolve?

We believe that the focal point for GRC practice needs to shift from the "risk category" perspective, that is a functional and departmental view of risk, and to align this with an enterprise-wide objectives-driven view. As the CSF-based model above demonstrates, the objectives-driven view is hierarchical and cross-functional in essence. In a business enterprise, upper nodes of the objectives hierarchy tend to be aligned with "survival" imperatives for the organization as a whole, followed by strategic objectives for the enterprise at Level 2, and so on. Hence, business objectives are cascaded or nested from top to bottom of the organization, across business lines and functional units.

> "The core of many of the problems banks face in managing risk across the enterprise is the manner in which they manage data; both structured data isolated in siloed databases and spreadsheets, and unstructured data in documents and text fields."

The current taxonomic or categorical view of risk in organizations is still important, as it represents a pooled area of valuable capabilities. However, once a risk has been identified, its importance or impact should be gauged by understanding where in the hierarchy of business objectives the impact of the risk lies. In addition, ownership of an objective should drive the focus on, or conception of, particular risks within the organization. It is also clear that where a risk that is known, controllable, but currently unmanaged, and which is identified as impacting on nodes in an upper hierarchy, should appear in a related continuous improvement log.

In this scheme of things, risk ratings are considered as objective measures. A risk with a high rating means that it has a singular impact on the related objective or node; a "medium" rating indicates that one or more risks in adjacent objective-risk nodes need to activate before impacting the upper level node; and a "low" rating means that all of the adjacent sibling nodes are required to activate before impacting on the objective-risk nodes at the next level above. Harmonization between risk categories should, however, be automatic, as risks that impact higher objectives rank higher than

those that rank lower.

We believe that risk classifications should, ideally, be system theory-focused. That said, some risks are "singular" in that if they crystalize they will impact the organization as a whole (e.g., regulatory fines for misconduct). Alternatively, some risks that occur lower in the risk hierarchy impact higher nodes (e.g., regulatory risk related to SOX, where, for example, it is assumed that managers take responsibility for the actions of staff). In addition, our proposed schema holds that uncontrolled risks at the bottom of the organization can have a compound impact at the top. Thus, risks need to be identified and controlled at greater levels of granularity.

This brings us to the fact that relevant decisions should be managed by the owners of business objectives. Risk mitigation should be dealt with on the intersection of objective – risk axis, as our model above indicates. The articulation of CSFs and related CRFs should help the design and implementation of related controls and to enable control testing. We have previously indicated that the model can also enable double-loop learning and enhanced decision-making. Thus, the application of our model will help to mitigate those risks caused by decision routines based on single-loop learning. However, we also note that other factors also influence risk decisions as decision makers often:

- Do not own the decision.
- Suffer from a raft of biases.
- Lack knowledge, skills, and capabilities.
- Have poor information and decision support.
- Are motivated to make the sub-optimal or incurred decision (e.g., through incentives).
- Are not motivated to take risk into consideration (i.e., are reckless).
- Make bad decisions deliberately (e.g., engage in misconduct or fraud).
- Make errors or are just negligent.

One of the key factors here is the value of information and IT-based system support to address what are basically information-related problems, be they information asymmetries or inability to access information within an organization.

10.3 The future of information systems support for enterprise risk management

The relative comfort the CFO faces in managing enterprise-wide financial data was noted, as was the importance of standards in communicating information and data. An integrative approach to identifying and categorizing objectives, risks, and controls with related critical factors and indicators provides a model for financial services organizations, as well as fintech or risktech vendors, to develop enterprise risk management systems that emulate financial information and enterprise systems.

It is clear that a riskless bank is a logical contradiction. Financial institutions take risks whenever they issue credit or trade in the markets. Such risks are financial and have both an upside and a downside; they are, therefore, undertaken in accordance with an institution's risk appetite. Non-financial risks have no upside and are all downside. Again, risk appetite and business impact of a risk event are the deciding factors. A bank, therefore, needs to be able to identify and distinguish between good and bad risks, in the context of financial and non-financial risks. These simple dichotomies could be used by a CRO and an enterprise risk team to frame their dialogue on risk with the business. Due to the siloed nature of the business and risk functions, as first and second lines of defense, communication is vital, and how such communication is framed is important. We have asserted, based on recent empirical evidence, that if the CRO and their team are not engaging in, or are not being included in, constructive dialogue with the business, then there are significant problems with risk culture in that institution.

We hold that the models we propose in this paper not only help address the good risk/bad risk problem, they also facilitate constructive dialogues at all levels within an organization. Hence, whether they are embedded in an enterprise risk management system or not, they are of material benefit to managers in creating the circumstances where such dialogues take place, with positive outcomes for the organization as a whole.

References

Abraham, S., and P. Cox, 2007, "Analysing the determinants of narrative risk information in UK FTSE 100 annual reports," British Accounting Review 39:3, 227-248

Alexander, K., 2006, "Corporate governance and banks: the role of regulation in reducing the principal-agent problem," Journal of Banking Regulation 7:1-2, 17-40

Altunbas, Y., S. Carbo, E. P. Gardener, and P. Molyneux, 2007, "Examining the relationships between capital, risk and efficiency in European banking," European Financial Management 13:1, 49-70

Argyris, C., 1976, "Single-loop and double-loop models in research on decision making," Administrative Science Quarterly 21:3, 363-375

Arwas, A., and K. Soleil, 2016, "Robo-Advice 2.0: the next generation," Journal of Financial Transformation 43, 30-36

Back, A., 2016, "Wells Fargo's questionable cross-selling strategy," Wall Street Journal, September 9, http://on.wsj.com/2uSSKNW

Bamberger, K. A., 2010, "Technologies of compliance: risk and regulation in a digital age," Texas Law Review 88, 670–739

Bennett, M., 2013, "The financial industry business ontology: best practice for big data," Journal of Banking Regulation 14:3-4, 3-4

BIS, 2016, "Standardized measurement approach for operational risk - consultative document," Bank for International Settlements, http://bit.ly/1nTDSai

Blunden, T., and J. Thirlwell, 2012, Mastering operational risk, Pearson Education Limited

Brandes, P., R. Dharwadkar, and D. Das, 2005, "Understanding the rise and fall of stock options compensation: taking principal–agent conflicts to the institutional (battle) field," Human Resource Management Review 15:2, 97-118

Bromiley, P., M. McShane, A. Nair, and E. Rustambekov, 2015, "Enterprise risk management: review, critique, and research directions," Long Range Planning 48:4, 265-276

Busenitz, L. W., and J. B. Barney, 1997, "Differences between entrepreneurs and managers in large organizations: biases and heuristics in strategic decision-making," Journal of business venturing 12:1, 9-30

Castelli, M., L. Manzoni, and A. Popovič, 2016, "An artificial intelligence system to predict quality of service in banking organizations," Computational Intelligence and Neuroscience 2016, article ID 9139380

CBC News, 2017, "TD Bank shares post worst day since 2009 after CBC story," March 11, http://bit.ly/2tFHcxV

CCP, 2015, "20 banks, five years and £252Bn in conduct costs," Conduct Costs Report 2015, CCP Research Foundation http://bit.ly/2tFIlFF

Chang, S. I., D. C. Yen, I. C. Chang, and D. Jan, 2014, "Internal control framework for a compliant ERP system," Information & Management 51:2, 187-205

Cocca, T. D., 2016, "Potential and limitations of virtual advice in wealth management," Journal of Financial Transformation 44, 45-57

Evans, L. T., and N. C. Quigley, 1995, "Shareholder liability regimes, principal-agent relationships, and banking industry performance," Journal of Law and Economics 38:2, 497-520

Gallagher, B. P., P. J. Case, R. C. Creel, S. Kushner, and R. C. Williams, 2005, "A taxonomy of operational risks," Software Engineering Institute, Carnegie Mellon University, http://bit.ly/2vWj06K

Grody, A. D., and P. J. Hughes, 2016, "Risk accounting-part 1: the risk data aggregation and risk reporting (BCBS 239) foundation of enterprise risk management (ERM) and risk governance," Journal of Risk Management in Financial Institutions 9:2, 130-146

ISO/IEC Guide 73:2002, Risk management – vocabulary – guidelines for use in standards, http://bit.ly/2uVi99s

Jarrow, R. A., 2008, "Operational risk," Journal of Banking & Finance 32:5, 870-879

Jessel, B., and T. Marshall, 2016, "Get bold with blockchain," Journal of Financial Transformation 43, 15-20

Kahneman, D., 2003, "Maps of bounded rationality: psychology for behavioral economics," American economic review 93:5, 1449-1475

Kaplan, R. S., and D. P. Norton, 1996, The balanced scorecard: translating strategy into action, Harvard Business Press

Kegan, R., and L. L. Lahey, 2001, "The real reason people won't change," Harvard Business Review 79:10, 84–92

Kim, S. D., 2014, "Characterization of unknown unknowns using separation principles in case study on Deepwater Horizon oil spill," Journal of Risk Research 20:1, 1-18

Liao, H. H., T. K. Chen, and C. W. Lu, 2009, "Bank credit risk and structural credit models: agency and information asymmetry perspectives," Journal of Banking & Finance 33:8, 1520-1530

Luft, J., and H. Ingham, 1961, "The Johari Window: a graphic model of awareness in interpersonal relations," Human Relations Training News 5:9, 6-7

McCreary, D. and A. Kelly, 2013, Making sense of NoSQL, Manning Publications

Mikes, A., 2008, "Chief risk officers at crunch time: compliance champions or business partners?" Journal of Risk Management in Financial Institutions 2:1, 7-25

Mikes, A., and R. S. Kaplan, 2015, "When one size doesn't fit all: evolving directions in the research and practice of enterprise risk management," Journal of Applied Corporate Finance 27:1, 37-40

Moosa, I., 2007, Operational risk management, Springer

Morris, J. J., 2011, "The impact of enterprise resource planning (ERP) systems on the effectiveness of internal controls over financial reporting," Journal of Information Systems 25:1, 129-157

Nelson, S. C., and P. J. Katzenstein, 2014, "Uncertainty, risk, and the financial crisis of 2008," International Organization 68:2, 361-392

Ormerod, P., 2007, Why most things fail: evolution, extinction and economics, John Wiley & Sons

Petraeus, D., 2015, Foreword: "Social science goes to war: the human terrain system in Iraq and Afghanistan," McFate, M., and J. H. Laurence (eds.), Oxford University Press

Rockart, J. F., 1979, "Chief executives define their own data needs," Harvard Business Review 57, March-April, 81-83

Rosemann, M., and M. Zur Muehlen, 2005, "Integrating risks in business process models," ACIS (Australasian Conferences on Information Systems) 2005 Proceedings 50, http://bit.ly/2uT7gFx

Sadgrove, K., 2016, The complete guide to business risk management, Routledge

Shojai, S., and G. Feiger, 2010, "Economists' hubris - the case of risk management," Journal of Financial Transformation 28, 25-35

Simon, H. A., 1955, "A behavioral model of rational choice," Quarterly Journal of Economics 69:1, 99-118

Soares, S., 2015, The Chief Data Officer handbook for data governance, MC Press LLC

Stanton, T. H., 2012, Why some firms thrive while others fail: governance and management lessons from the crisis, Oxford University Press

Stulz, R. M., 2015, "Risk-taking and risk management by banks," Journal of Applied Corporate Finance 27:1, 8-18

Taleb, N. N., 2005, The black swan: why don't we learn that we don't learn, Random House

Taleb, N. N., 2012, Antifragile: things that gain from disorder, volume 3, Random House

Tett, G., 2010, "Silos and silences. Why so few people spotted the problems in complex credit and what that implies for the future," Financial stability review 14, 121-129

WEF, 2016, "The role of financial services in society: understanding the impact of technology-enabled innovation on financial stability," World Economic Forum, http://bit.ly/1qCGJWN

Zuboff, S., 1991, Informate the enterprise: an agenda for the twenty-first century," National Forum 71:3, 3-7

Reconciliations: Five trends shaping the future landscape

ARIF KHAN | Principal Consultant, Capco

ABSTRACT

Reconciliations are found throughout the financial services industry. In an increasingly complex world, with stricter regulatory requirements, reconciliations are applied heavily, and contribute significantly to the cost of doing business for financial institutions (FIs). This paper aims to explore some of the key emerging trends in the world of reconciliations. It looks at how cutting-edge technology, such as blockchain, machine learning, and robotic process automation (RPA), combined with the move to reconciliation managed services, are defining the reconciliation model for the FIs of tomorrow.

1. INTRODUCTION

Financial institutions have long relied on reconciliations as a key control to ensure accurate data. Reconciliations are not only essential in accounting and in the finance areas, but are heavily used across the capital market space. Any operational department in a financial institution (FI) will have many such processes, typically reconciling with clients, prime brokers, and external exchanges. Reconciliations were introduced as key controls in operational processes, yet seemed to have spawned beyond this. Increased regulatory scrutiny, larger amounts of data, and increasingly complex financial products have led to operational departments having to operate hundreds of reconciliations daily.

The macro picture for FIs is an environment of falling revenues, increased cost of business due to regulation, and the constant need to "do more with less." Cost reduction is more of a focus than ever. At the same time as this pressure is being applied there are technological advances that are claiming they will change the way we do business for ever.

This paper looks at five key trends shaping the world of reconciliations, ranging from industry-wide utilities to artificial intelligence (AI) technologies that will automate key areas of the processing.

2. BACKGROUND

Reconciliations are essentially checks to ensure that two or more data sources agree with each other. They are typically performed between two points in a business process. Some examples of the types of reconciliation typically found in a FI are:

- **FOBO (front office to back office)** – risk system reconciled to books and records platform.
- **Exchange** – FI's trade, position, and cash records reconciled to clearing houses records.
- **Nostro** – payments made and received reconciled between books and records and Nostro bank account.
- **General ledger** – reconciling the general ledger to the relevant sub-ledger.
- **BOBO (back office to back office)** – reconciling back office data with another source from the back office.
- **Inter-system** – data integrity and completeness check between two systems.
- **Trading: total equity** – a combined reconciliation of trades, position, and cash between a central clearing party and a clearing broker, or the broker and a customer.

This paper explores the following five areas that have been identified as trends defining the reconciliation landscape of tomorrow:

1. Automation of manual reconciliations via self-service tooling
2. Elimination of intersystem reconciliations
3. Blockchain and distributed ledger technology
4. Outsourcing reconciliations to industry utilities
5. RPA and machine learning

This paper will look at the cause of these underlying trends, and explore how each is changing the market offerings around reconciliations.

3. AUTOMATION OF MANUAL RECONCILIATIONS VIA SELF-SERVICE TOOLING

The first area to look at is the problems associated with slow on-boarding times for new reconciliation processes, and how this problem is being tackled by the fintech world.

3.1 Long onboarding time

Aite Research group concluded that it takes on average 64 days to set up a single new reconciliation.[1] On-boarding a reconciliation on to SmartStream's TLM platform, a leading vendor known by a majority of FIs,[2] takes between 22 days and six months.[3] This causes an immediate issue for business and operational units requiring rapid turnaround of changes to reconciliations. These units are under increasing pressure to not only address changing regulatory and client needs, but to also fix bugs in the existing reconciliations themselves. There is a valid discussion to be had around the cause of this time frame, not least the huge variance. It is not necessarily correct to attribute a long on-boarding time solely to the vendor system being employed to perform the reconciliations. Asked about the six-month on-boarding timeframe for new reconciliation on to TLM, often quoted by operational users, Rocky Martinez, CTO of SmartStream highlighted that "It's not actually

[1] Aite Group LLC, Feb 2016, "Reconciliation trends in 2016: regulation and nervous recs," 19
[2] Aite Group LLC, April 2014, "Reconciliation Technology Solutions in 2014: recs get ready to rumble …," 25
[3] Ibid, 27

the TLM product itself, it's the data received. When we receive data from the customer it needs to be cleaned then fed into the various production cycles. It's actually a pretty complex operation but it's a unique part of the service we provide."[4]

Data preparation is a key part of any reconciliation, with users often combining the data sourcing effort with the building of a reconciliation, when obtaining an estimate. A six-month turn around would also typically incorporate a period of "user acceptance testing" (UAT), which is often set at one-to-two months by the internal policy of FIs.

3.2 Firm-wide reconciliation groups

There is another important factor to consider when understanding the long on-boarding times associated with these platform, and that is the creation of central firm-wide functions within FIs. These groups are responsible for the on-boarding of new reconciliations and management of reconciliations output and platforms. In a move to create these teams within their companies and to obtain economies of scale, several FIs have created centralized technical and operational expertise around reconciliations in a single low-cost location. These are often referred to as "centers of excellence" (CoEs). While moving to this model with CoEs reduces cost, centralizes governance, and co-locates those working on reconciliations, it has one serious downside; namely, that it creates a bottleneck for any requests for changes to these platforms.[5]

3.3 MS Excel- and Access-based solutions

Over time, the bottleneck from CoEs, along with the prospect of a six-month turnaround time, leads to teams building tactical solutions. The tools many turn to are Microsoft Excel and Microsoft Access, software typically available to all users in the institution, and ones operational users work with daily. These tools allow operational users to build their solutions in a few days and quickly apply them to their operational procedures. Unsurprisingly, these tactical approaches end up becoming embedded in procedures and lead to several key problems,[6] such as manual processes using up operational capacity each day, working against cost cutting initiatives, solutions not being scalable, reliance on a single user with knowledge of the control, hence increasing the risk of fraud, and lack of audit details, or any metrics, around these solutions.

3.4 "Self-service" tooling

This is the area targeted by reconciliations solutions that offer a "self-service" capability to operational users, allowing non-technical users to build reconciliations directly and in a short period. These solutions are scalable platforms with features such as audit trail and version control. One such vendor is DUCO.

DUCO's CEO, Christian Nentwich, explains their market focus in the following words: "What we really go after is all of the work that banks still do manually. There are a lot of people armed with spreadsheets and highlighter pens. All the labor arbitrage is already done, so they may be sitting in some offshore locations, but they are still doing it manually. At the end of the day comparing data is not a job for humans."[7]

DUCO aims to empower non-technical users to directly build and run reconciliations. Their DUCO Cube solution utilizes technology in "natural language processing" (NLP), along with an intuitive "user interface" (UI), to enable configuration and set up time for new reconciliations to be greatly reduced. The formula appears to be gaining traction with company revenues up 120% in the last year and the company growing rapidly. DUCO's technology leads to the ability to rapidly set up reconciliations, with an average set up time of 2.4 days versus the industry average of 64 days.[8]

There are other solutions on the market that have a similar approach, targeting those reconciliations done manually or via Excel macros. Once such solution is RecsHub from the vendor Xceptor. The solution also utilizes a rules-based configuration that allows users to "define and manage their reconciliations processes, without having to rely on IT support."[9] While DUCO aims for an NLP-based approach to make rules easy to configure, Xceptor RecsHub uses the paradigm of MS Excel using the same names and syntax for common functions. Operational users can configure rules to process data and perform the matching directly via the UI and not require IT intervention.

[4] Rocky Martinez – SmartStream CTO, 4th September 2017, personal interview
[5] Paul Clapis, Vice President, Engineering and Architecture, Reconciliation, Institutional and Wholesale, FIS, 26th September, personal interview
[6] Ibid, 18
[7] Christian Nentwich – DUCO CEO, August 22nd, 2017, personal interview
[8] Keith Whelan – DUCO Managing Director, EMEA, August 2017, "Reconciliations: from boring necessity to key business function."
[9] Xceptor reconciliation hub, http://bit.ly/2x941Xq

3.5 What is the target operating model?

These self-service tools solve an immediate problem and they do so rapidly, as per their design. Operational teams benefit from some "quick wins," as they are able to avoid a lengthy technology book of work with an optimized reconciliation process. It remains to be seen, however, whether this model of self-service lead by operations is sustainable within large organizations. Do operational teams want to be responsible for the maintenance and upgrades to any reconciliations, along with their tasks of day-to-day processing? In a decentralized model, such as this one, where are the governance and controls around the process to avoid duplications and the creation of reconciliations that may not be needed in the first place? Introducing additional platforms for any business process leads to teams having to manage split-processes and lack of a single combined view of a given function for management. The running of reconciliations, the workflow around management of breaks, and the dashboards and management information views these tools provide are immediately more complex with multiple tools.

The "silver bullet" of self-service reconciliations appears to certainly solve one problem but, unless appropriate governance and processes are put in place, it does in fact create new issues.

Despite these issues, there is a clear argument to be made that having these reconciliations on a platform, such as DUCO or RecsHub, is a step forward from having them done completely manually on a spreadsheet. Managers may not be able to easily get a single overview of all of their reconciliations, but they are getting far greater control and audit capability than when the process was manual.

3.6 Temporary reconciliations

Another everyday use case for reconciliations that aligns perfectly to the self-service tools is the area of temporary reconciliations. These throw-away reconciliations are useful to add control during an operation such as a system upgrade or migration. Teams require a reconciliation process to be in place to ascertain the successful completion of the activity but will no longer need the reconciliation after this. The only economically viable solution for this type of reconciliation is something that can be set up quickly and easily without requiring IT involvement.

In the current landscape, there is sufficient justification for self-service reconciliation systems to be employed by FIs. This split model is only justified, however, because the better-established platforms have failed to make their solutions fast enough for on-boarding new reconciliations. If the more established reconciliation platforms can solve this problem and make their products more agile, and open to "self-service," then they can once again claim to offer "one-stop-shops" for reconciliations. We will return to this topic in the section on machine learning and look at how one such vendor, FIS, has responded to this situation.

> "The introduction of a true distributed ledger means that multiple reconciliations are avoided as the accuracy of the single, shared representation of the contract is agreed upon via a consensus algorithm."

4. ELIMINATION OF INTERSYSTEM RECONCILIATIONS

The main use cases for reconciliations at "sell-side" firms are internal (69%) and intersystem (57%) reconciliations.[10] These reconciliations are borne out of the complex IT architecture often found in back offices of large FIs. Typically, multiple systems contain data relating to trades, positions, and balances at different stages of the trade lifecycle, and are reconciled to ensure that they are aligned.

In 2016, when announcing their Strategy 2020 vision to investors and the wider public, Deutsche Bank revealed that they have over 1,000 intersystem reconciliations.[11] At the same time as unveiling this figure, they also announced ambitious targets to reduce these by 70% to around 300 by 2020. If reconciliations are essential controls in processes, what is the approach for removing so many reconciliations? The fact that they can be reduced by 70% illustrates that there is a level of redundancy here. Below we look at the options that exist for replacing intersystem reconciliations.

4.1 Mis-use of reconciliations?

"Reconciliations are borne out of an insecurity around data process," explains an enterprise architect at a Tier-1 Global Investment Bank. "Whenever data crosses a

[10] Ibid, 9
[11] Deutsche Bank, 2016, "Deutsche Bank annual press conference," http://bit.ly/2g3qv6D

boundary, system or organizational, there is a demand for a reconciliation. In large financial institutions, there are a lot of these boundaries."[12]

This approach to data validation naturally leads to very large numbers of intersystem reconciliations, which quickly become embedded into operational procedures. To reduce the number of these checks, a holistic review of the end-to-end process is required. "We actually talk to many banks about this." explains Christian Nentwich, CEO and co-founder of DUCO. "The ones that are more active investigate which of these manual controls they actually need. There is definitely a thread running here where people say that internal system reconciliations mask systemic issues that shouldn't occur in the first place."[13]

The back office of investment banks is one area where large numbers of reconciliations are typically found. Many of the core back office systems are based on outdated technology and batch processing. This architectural landscape is another key factor in driving data validation towards reconciliations, invariably run post batch on a T+1 basis. There are solutions to this problem. The first solution has been around for many years, and is part of a well-established industry-wide trend of moving from "T+1" data processing to a "T0" world.

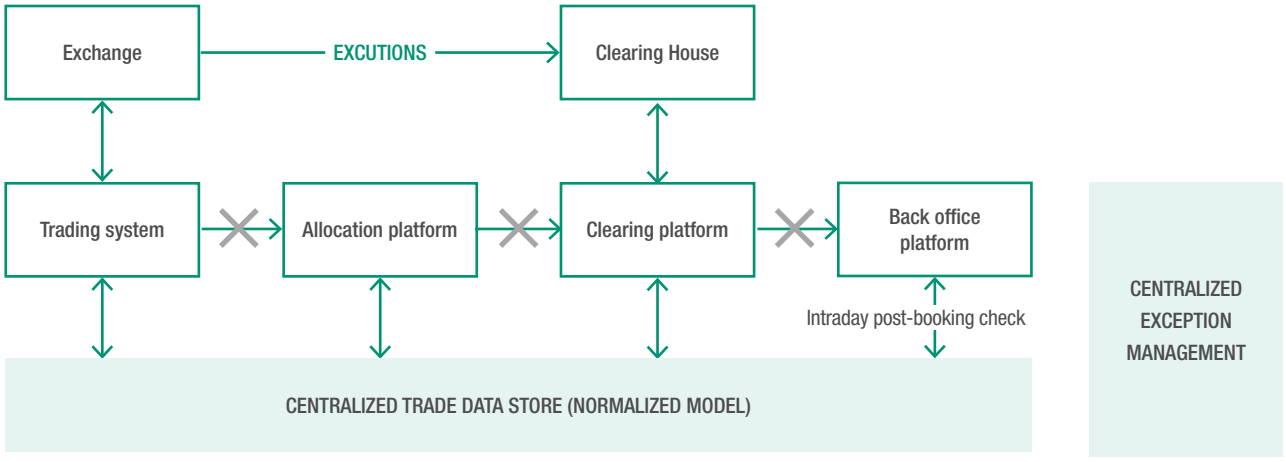
4.2 Real-time exception-based processes

The concept of a centralized trade data store is one that is well known in the world of capital markets. As regulation tightened around regulatory reporting, many firms adopted such solutions, centralizing their trade data in a single repository. This same approach of centralizing data can directly help with one subset of intersystem reconciliations, which are those performed during the trade lifecycle.

Figure 1 shows an example of an architectural blueprint for moving from point-to-point flow to a centralized model, where a single version of the trade is updated, and enriched during the trade lifecycle.

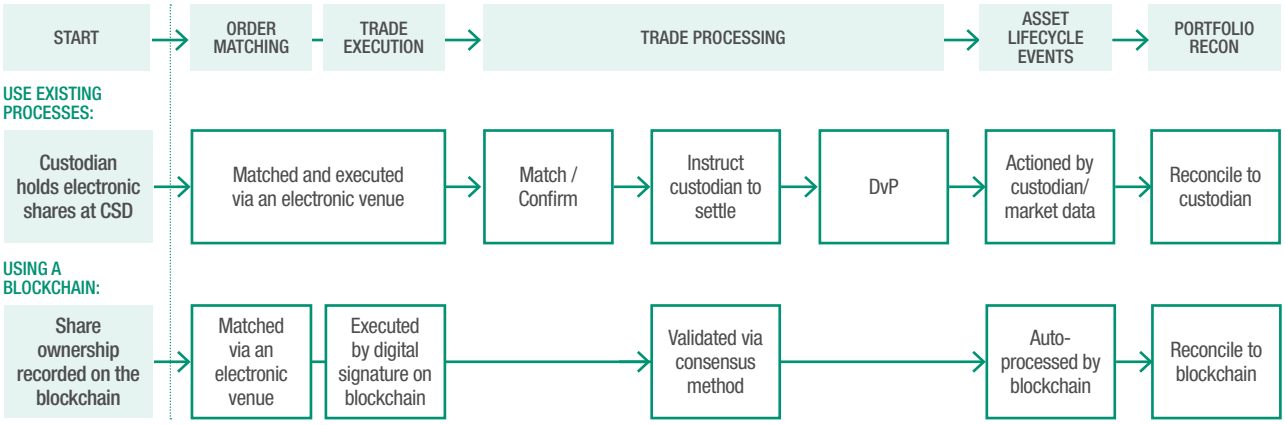
This model is combined with centralized real-time exception management, and a single user interface (UI) showing the status of the trade. The final part of the trade flow is booking the trade into the books and record platform. The books and record systems are often platforms that are decades old, and not built for real-time messaged based processing (for example, ION's RANsys listed derivatives back office processing platform). In this situation, using message queues is not always possible. In Figure 1, a proposed solution for this challenge is an intraday post-booking check.[14] This approach moves the validation of the booking to a T0 process, rather than something that would need to be picked up in a T+1 post trade reconciliation.

Figure 1: Example flow through trade lifecycle – listed derivatives



[12] Enterprise Architect Tier-1 Global Bank, September 14th, 2017, personal interview
[13] Christian Nentwich – DUCO CEO, August 22nd, 2017, personal interview
[14] Christian Nentwich – DUCO CEO, August 22nd, 2017, personal interview – "If your STP is perfect why should you need an intersystem reconciliation? At least if you do perform this reconciliation, do it in real-time instead of these big batch processes."

Figure 2: Equities transaction trade lifecycle – current process alongside possible blockchain process



Source: Innovate Finance blockchain DLT and the Capital Markets Journey – Oct 2016

This not only eliminates the need for an intersystem reconciliation, but also helps errors to be captured and corrected on T0, before they impact any calculations or downstream processes (e.g., client margin calls).

Moving to this form of architecture takes time and intersystem reconciliations may need to be in place for a temporary period while the system is tested and operational teams and managers gain confidence in the flow. Re-engineering trade flows in this way, and moving to more real-time data validation will allow FIs to eliminate many intersystem reconciliations. While re-engineering existing solutions in this way may be costly, at a minimum FIs should ensure that any new applications being rolled out conform to this real-time data validation pattern and additional intersystem reconciliation are not put in place.

5. BLOCKCHAIN AND DISTRIBUTED LEDGER TECHNOLOGY (DLT)

A more cutting-edge technological approach to reducing the number of reconciliations is to have a single immutable representation of the data. A blockchain is a type of distributed ledger, comprised of unchangeable blocks of digitally recorded data. Each link in the blockchain includes a check to validate and ensure the data has not been altered. Crucially, there exists a single, shared view of the data, rather than multiple versions requiring reconciliation. The accuracy of the data on the blockchain is verified through consensus validation, with a single audit log showing the chain of events.

There is scope for this technology to dramatically alter the landscape of how trades are executed, processed through their lifecycle, and settled. This overhaul of the way financial markets operate would impact reconciliations across many areas. If the solution was shared across multiple FIs, exchanges, clearing houses, and regulators, then DLT can have a revolutionary impact. Figure 2 maps out what a theoretical future blockchain based solution for processing an equities transaction may look like in comparison to the current process.

The elimination of reconciliations is regularly cited as one of the key tangible benefits of the industry adopting this form of technology and architecture. As we have seen, reconciliations are a growing problem and FIs are constantly on the lookout for ways to reduce the costs associated with them. Currently, financial markets operate "based on the logic of 'consensus-by-reconciliation'" [Morini (2017)].[15] The only way confidence is established in the details being accurate is if both counterparties have the same records in their respective systems. Getting to this agreement spans across multiple business processes, such as "confirmation, affirmation, communication to central bodies."[16] The introduction of a true distributed ledger means that multiple reconciliations are avoided as the accuracy of the single, shared representation of the contract is agreed upon via a consensus algorithm. This approach moves the paradigm from a "consensus-by-reconciliation" to a "distributed ledger" model.[17]

[15] Morini, M., 2017, "From 'blockchain hype' to a real business case for financial markets," Journal of Financial Transformation 45, 30-40
[16] Ibid, 32
[17] Ibid, 38

5.1 Smart contracts

These techniques can be applied to a deal consisting of multiple payments, like a bond, through the concept of a "smart contract." These contracts aim to model financial contracts in self-contained modules of programming code. The agreement on the terms of the smart contract take place at the start of the trade lifecycle, and from then on the relevant counterparties agree to refer to the single version of the deal on the distributed ledger. Combing DLT and smart contracts provides the basic building blocks for creating a model of the financial markets that operates in a very different way to what we have today.

Is this technology going to make the world of reconciliations obsolete soon? While there is a growing view that this is the future model for how things should work, there is a nagging feeling that the speed of change will not be as fast as many would hope for. The world of financial services and large FIs are not known for rapid adoption of new trends, and fast establishment of new standards or technologies. In addition, there are many key areas of technology and defined standards that would need to be in place for such a solution to get off the ground:

1. **Privacy:** ensuring the data security around financial data for multiple FIs, which is very sensitive and in a highly-regulated area.
2. **Scalability:** there are no proven solutions combining DLT with smart contracts at the scale that would be required for this area.
3. **Definition of standards:** standards for the definition of smart contracts would need to be agreed and formulated, with strong opinion already voiced that fpML would not be fit for purpose

Christian Nentwich summarizes this when talking about DUCO's strategy: "On the bet of where you make your money in the next five years; is it doing what we do or is it blockchain? In the near term, I bet on what we do."[18]

5.2 Distributed ledger – internal

The chances of this sort of major fundamental change to the core infrastructure of the capital markets within the next five years are slim. What is, however, far more plausible and achievable within that time frame is the introduction of a distributed ledger within a given organization. As the data passes through the trade lifecycle, each system in the chain is referring to a single shared representation of the data, rather than having its own copy. This change would lead to far less of an impact in the financial markets than a distributed ledger across many external parties, but it would be one practical option for eliminating multiple copies of data within FIs, and, thus, multiple reconciliations. Internal systems in the FIs are often multi-instance and geographically spread. A consensus-based model would bring benefit in to this environment. Is this approach to the issue new, or is it just re-packaging the centralized data store?

Centralized data stores are not a new concept but the focus and buzz around blockchain could become a catalyst to firms adopting better architectures. As the amount of data stored and analyzed grows, there is an increased focus on firm-wide data quality, and data lineage, which will get these topics on to the agenda of CIOs. DLT also has some key differences from a centralized data store that need to be understood.[19] Firstly, DLT increases fault tolerance and avoids a "single point of failure." Secondly, it avoids centralized operating risk. Finally, it avoids risk and accusation of central data store owner of manipulating the data.

In conclusion, DLT has the potential to eliminate large numbers of reconciliations[20], but the market is many years away from having a working solution. It remains to be seen if such a solution will materialize, although it should be noted that industry committees are formed already and proposing specific work streams in this space.[21] The introduction of this technology into the wider market place, however, is an excellent opportunity for FIs to re-evaluate and re-engineer their solutions. Adopting such an approach, even within a single FI, would allow for the reduction in intersystem reconciliations.

6. OUTSOURCING RECONCILIATIONS TO INDUSTRY UTILITIES

Another way to eliminate the burden of setting up, running, and maintaining reconciliations, is to outsource them. As FIs become more open to allowing their data to be shared with vendors, and hosted on systems outside their physical core network, new options are opened for managing reconciliations. Aite's surveys showed that the number of respondents having "no interest in

[18] Christian Nentwich – DUCO CEO, August 22nd, 2017, personal interview
[19] Morini (2017)
[20] Innovate Finance, 2016, "Blockchain, DLT and the capital markets journey: navigating the regulatory and legal landscape," October, 32
[21] ISDA MITOC, September 2017, "Data and process standards," 2

managed reconciliation services" dropped rapidly from 57% in 2013 to just 17% in 2015.[22]

One such vendor offering this is SmartStream and their CTO explained their entry in to this space: "As we've seen the mass adoption of outsourced reconciliations solutions it just makes sense to use a provider, like SmartStream, that can supply the only purpose-built solution that is totally agnostic and works at an enterprise level with many, if not all, complimentary services the customer may need."

6.1 Differing levels of outsourcing

There are several different models for the outsourcing of reconciliations, with each iteration giving a little more of the process over to the vendor. This is illustrated in Figure 3.

The appetite for the different models was gauged as part of Aite's research, which found that the most popular offering was the "fully managed service," with 52% of respondents expressing interest in this model. The "partial service" had 17%, and "hosted service" and "full outsourcing" had 8% and 7%, respectively.[23]

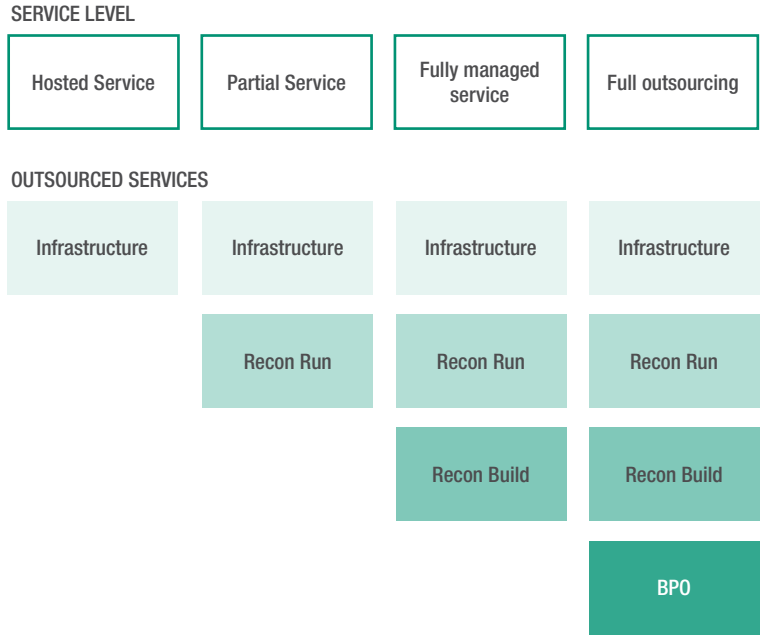
The adoption of even a basic hosted service does provide some tangible benefits such as:

- **System administration:** banks no longer require in-house teams responsible for maintaining the IT hardware and keeping systems up and running. For the utility, this can be a service offered to multiple clients using a shared pool of resources, providing some efficiency.
- **Harmonize versioning:** all platforms can be upgraded to the latest format, having a consistent edition across the infrastructure. This reduces complexity and unlocks the features of these latest versions.
- **Latest technology releases:** the vendor can apply and gain benefit from tooling that has either not yet been released to the market, or is a module the FI had not previously taken out. These new modules and upgrades can be directly applied across the client base.

6.2 Full outsourcing

It is not surprising to see that the least popular option in the Aite survey, conducted in 2015, was the "full outsourcing" model. This model is the newest and the most drastic in terms of the level of responsibility handed over to a third-party. Could moving to this model really be an effective way for FIs to reduce some of the cost and headache of maintaining and running these processes in house day-to-day?

Figure 3: Differing levels of outsourcing and the corresponding services provided by the vendor



The utility model is gaining popularity with several vendors and consultancies announcing such solutions in the past two to three years. To obtain the benefits of such a model, and for the utilities to be successful business operations generating profit for the owners, increasing efficiency must be top of the agenda. One market utility that appears to be gaining critical mass is the FIS Derivatives Utility. FIS launched this utility in 2015, when SunGard (now owned by FIS) partnered with Barclays as their anchor client. SunGard took not only the hosting of the post-trade processing, but also the management of the operational services around the technology, resulting in a complete outsourcing offering to the market.

Unsurprisingly the views from the industry vary depending on which part of the FI landscape you are in.

One middle office risk and control manager from a Tier-1 Global Investment bank questions the business benefit of moving to the model offered by FIS. The officer initially questions if it is "really a utility or just moving teams and systems?" He further highlights the lack of a "standard model" and questions if "banks [are] just

[22] Ibid, 27-28
[23] Ibid

doing this as a short-term way to reduce their bottom line?"[24] This is a valid question to be raised, as in some cases, such as Barclays adoption of the FIS Derivatives Utility, employees from the FI were transferred over to the provider to then provide the same service they were providing previously.[25] Is this utility a significant step for the industry, or is it just a "lift-and-shift" of people and technology?

Christian Nentwich of DUCO also questions the first step of this model, describing it as "pretty underwhelming" given that the cost savings are around 20%.[26] He believes that to really unlock the potential of this type of offering, the focus needs to move to the "changes and improvements [that] are required to achieve savings of 50-60% rather than these marginal gains."[27]

Richard Chapman, a VP in Strategy and Business Development, Reconciliations for FIS, explains how step one of the process is indeed purely about moving the processes "as-is" out of the FI and into the utility with virtually no changes. Soon after this, however, he points out that you "quickly start to identify, in a much clearer fashion, bottlenecks and pain-points."[28] This starts the process of optimizing the processes and improving the efficiency of the utility. In the case of the FIS Derivatives Utility, Richard explains their approach for unlocking the real value of the utility: "The focus with a utility is on realizing economies of scale; how can you get efficiencies so you can reduce cost and increase market adoption quickly? This naturally now leads in to Artificial Intelligence and, in particular, machine learning."[29]

7. RPA AND MACHINE LEARNING

The drive for more efficiency gains greater purpose with the introduction of a utility. Today, it is the latest technology trends that are enabling faster reconciliation set-up and automation around resolving the breaks. The utilities have strong business justification for investing heavily in advanced technology, as high-levels of automation are fundamental to their business model.

While there is a lot of talk and hype about the potential for applying artificial intelligence techniques to the area of reconciliations, few practical examples of this exist outside innovation labs or proof of concept builds.[30] One solution that has been released to the market place, and does utilize this technology, is Intellimatch Accelerator.[31] Head of Product Management for Reconciliations at FIS, Michael Maggio, explained the approach of focusing AI technology where FIS were seeing the "biggest pain points" for their clients and "in the broader market."[32] Pain was observed in both "reconciliation construction and reconciliation execution."[33]

7.1 Automation: on-boarding reconciliations

FIS claim that the Accelerator product brings the time required to create a new reconciliation down from 45 hours to one hour.[34] The marketing material for the tool talks of the system "automating" the on-boarding of new reconciliations, but what techniques are being employed here and is this possible?

There are three main techniques employed in the setting up of the reconciliation. Each technique is employed to provide the user with the information to streamline the setup of the reconciliation.[35]

- **Heuristic techniques** – automate the analysis and mapping of data fields based on previous data.
- **Match data quality** – present the user with details of match quality and match rates for potential matches.
- **Direct feed** – this on-boarding tool can feed rules directly in to the existing Intellimatch platform and is not a third-party analysis tool sitting outside the process.

As discussed, empowering operational and business users through self-service tooling resolves one of the key bottlenecks in the setup of reconciliations. Combining this with machine learning techniques, to detect the quality and integrity of data and to suggest potential matches, speeds up the process even further.

Work is already underway on additional features of the tool to further improve the experience for users, and reduce on-boarding time. Michael Maggio explains the latest feature of this platform, which is "to allow clients to focus purely on defining their specific reconciliation

[24] Operational Risk and Control Manager Tier-1 Global Bank, September 13th, 2017, personal interview
[25] SunGard, Press Release, 2015, "SunGard launches industry utility to transform derivatives clearing processing globally," http://bit.ly/2yjgvB2
[26] Christian Nentwich – DUCO CEO, August 22nd, 2017, personal interview
[27] Ibid
[28] Richard Chapman, Vice President Strategy and Business Development, Reconciliation, Institutional and Wholesale, FIS, 26th September, personal interview
[29] Ibid
[30] Christian Nentwich – DUCO CEO, August 22nd, 2017, personal interview – DUCO have performed research into completely autonomous set ups. Christian stated they had done some work in this area and planned to do more in the future.
[31] FIS Global, 2017, "Intellimatch Accelerator – reconciliation | automating the creation and refinement of reconciliations," http://bit.ly/2yPYLu1
[32] Michael Maggio, Vice President, Head of Product Management, Reconciliation, Institutional and Wholesale, FIS, September 26th, 2017, personal interview
[33] Ibid
[34] FIS Global (2017)
[35] Paul Clapis, Vice President, Engineering and Architecture, Reconciliation, Institutional and Wholesale, FIS, September 26th, personal interview

business process, and allow our Artificial Intelligence engine to do the rest."[36] This will empower the end-users by graphically representing the business process they are looking to reconcile, and using this as a further input to the AI engine for automating the setup.

Although it appears that the Intellimatch Accelerator is the first tool to market bringing together all these concepts, other vendors are actively working in this space. Both DUCO and SmartStream cited automation around the reconciliation setup process as an area of research they were involved in.[37] The AI techniques mentioned are a good fit for the automation of these reconciliation process setups. The number of vendors and offerings that utilize these techniques will only increase, and in turn the time taken to set up reconciliations will continue to decrease.

7.2 Automation: rule-tuning

One area not highlighted to date in this paper is the concept of the degradation of matching fidelity over time with a given reconciliation process. A reconciliation running today with a high matching rate can run tomorrow with a lower rate of match quality, as the rules are outdated or due to changes in the data source that have not been reflected in the business logic. The values being highlighted are in fact false positives, rather than genuine business breaks that require attention. This problem is serious enough for it to be another focus for FIS. Once more, it is machine learning that is being utilized to solve this problem. "We keep track of what staff are doing" explains Michael Maggio, "to manually correct issues caused by degradation, and then use machine learning techniques to automatically replicate this activity."[38]

Through this product, FIS have been able to take machine learning and AI technology and apply it directly to these two problem areas of reconciliations. This will speed up the onboarding of new reconciliations, and help to keep them running effectively. What they do not help with, however, is the manual process of handling the genuine breaks that are highlighted by the reconciliations. Workflow features are very common in reconciliation solutions, allowing different breaks in different reconciliations to be assigned to different teams or individuals, yet resolving these breaks is still a manual process.

In the final section of this paper it is this problem to which we turn our attention. Can this manual process of resolving breaks be something we automate using RPA?

7.3 Automation: break resolution via RPA

RPA is another of the buzzwords of the moment within

[36] Michael Maggio
[37] Rocky Martinez – SmartStream CTO, September 4th 2017, personal interview, "We're currently in discussions regarding the use of our data dictionary and the automation some of the previously manual processes."
[38] Michael Maggio

the technology innovation space. RPA is a software robot that simulates human actions through user interfaces. These robots are setup to perform the same processes that humans currently perform. While RPA may conjure up the same imagery of autonomous machines as machine learning there is an important difference. Paul Clapis, VP Engineering and Architecture – Reconciliations for FIS, explains: "It [RPA] is an interesting contrast to the machine learning [we use]. The strength of RPA is in automation of tasks that are distributed across multiple systems, but that are highly repeatable."[39]

Through collaboration with their RPA function based in Pune, the FIS team are in the process of building out solutions to automatically resolve breaks that occur. There are three AI techniques that are being used in unison to provide the required functionality in this space[40]:

(1) **Heuristic techniques** – looking at how humans make decisions and applying some basic logic via rules.

(2) **Classification techniques** – taking the actions that users perform as a training set of data. Once trained on these examples the system can then make predictions about appropriate remediation. As more examples feed through the system, and it is retrained with more data, the quality improves.

(3) **Clustering techniques** – identifying patterns in actions taken around resolving breaks that users themselves had not identified.

7.4 Remediation activity

These AI techniques allow the system to be able to predict what the appropriate action is to resolve the break, and it is RPA that will allow these actions to be performed automatically. An example of such an action would be a robot connecting to an upstream reference data source and inserting a missing product ISIN in that system, in response to a reconciliation break on ISIN.

This combination of machine learning and RPA to automate key problem areas of the setup and management of reconciliation is a powerful one. The timing of this technology coming of age is ideal for reconciliation utilities. These technologies can provide the automation they require to gain the efficiencies they need to make their models profitable. Even better news for the utilities is that this technology is available today and at least one vendor is already having success applying these methods.

8. CONCLUSION

In conclusion, it is clear to see that the reconciliation landscape is evolving, and there are many different drivers behind this. The push for reduction in cost is forcing FIs to look at outsourcing models that were not even offered in the market a few years ago. Utilities are here to stay and more companies are actively pitching to FIs in this space. As these utilities strive for better efficiency, it is technology, such as machine learning and RPA, that is proving a key enabler for them to achieve economies of scale through automation.

The impact of blockchain remains to be seen. There are several key obstacles for its adoption as a multi-institution platform, where it would provide the biggest benefit and disruption to the world of reconciliations. The technology may bring some benefit for FIs by re-engineering their internal processes as an alternative to a pure centralized data store, moving from a "consensus-by-reconciliation" model to automated algorithms already used in the world of Bitcoin.[42]

These technology advancements for FIs are good news and very timely. FIs need to re-define their reconciliations strategy in response to these new market offerings and start to reap the rewards of these cutting-edge technological developments.[41]

[39] Paul Clapis, Vice President, Engineering and Architecture, Reconciliation, Institutional and Wholesale, FIS, September 26th, 2017, personal interview
[40] Ibid
[41] Ibid
[42] Coindesk, 2014, "How bitcoin mining works," December, http://bit.ly/2vGaOdz

Thank you and goodbye – ending customer relationships and its significance

DAVID LIM | Senior Consultant, Capco

ABSTRACT

In today's world of increased regulation, scrutiny, and cost cutting, financial institutions are under pressure to ensure their client portfolios have the appropriate level of risk and return. Financial services firms need to be clear on their client selection and exit management strategies to drive revenue growth through new and existing relationships, whilst in parallel, ensuring risky and uneconomical clients are managed accordingly.

Managing risky or uneconomical clients is a challenge that many institutions face, and most are fully cognizant of just how reputationally damaging it can be when not done in the most client sensitive way.

Why would or should client relationships be terminated in the first place? What benefit can these organizations gain from ending relationships with their clients? How can this be done in a sensitive way?

This article looks at key considerations for financial institutions and the challenge of terminating relationships with existing clients. Moreover, it explains why ending relationships with existing clients should be an important part of an organization's agenda and how they can position themselves to do this in a sensitive and sensible manner. The article additionally explores how terminating relationships can best be done to reduce the reputational risks presented to them.

1. INTRODUCTION

In today's world of increased regulation, scrutiny, and higher operational costs, financial institutions are under pressure to ensure that their client selection and exit management strategies help drive revenue through developing new businesses, onboarding, and building strong relationships with customers, while at the same ensuring risky and uneconomical clients are managed accordingly.

This is not an easy task, as informing clients that their business is no longer wanted is often met with confusion, frustration, and resentment. Clients are more likely to understand and accept when a financial institution refuses to on-board them as a client, but are less forgiving if, after years of what they feel is valuable business, they are told that their business and relationship is no longer wanted.

This is a challenge that many institutions face, and most are fully cognizant of just how reputationally damaging it can be when not done in the most client sensitive way.

Why would or should client relationships be terminated in the first place? What benefit can these organizations gain from ending relationships with their clients? And, how can this be done in a sensitive way?

This article will look at the key considerations for organizations and the challenge of terminating relationships with existing clients. Moreover, it explains why ending relationships with existing clients should be an integral part of an organization's agenda and how they can position themselves to do this in a sensitive and sensible manner. The article further explores how terminating relationships can be done in such a way so as to minimize the reputational risks associated with such a decision.

2. WHAT IS CLIENT SELECTION AND EXIT MANAGEMENT?

Client selection is the strategy and approach that defines which clients a firm should engage with (on-board or build new relationships with), retain (continue to develop relationships with existing customers), or exit (ending the relationship with the existing client) based on a risk assessment of the client against the firm's risk appetite range or profitability criteria. Client size, the products the institution has, location, credit rating, and risk rating, are just some of the key aspects that should be incorporated with an organization's client selection strategy.

Client "exit management" is the governance and execution of off boarding clients. As part of this process, financial institutions must ensure that all accounts, products, services, and ultimately, relationship, associated with the client is closed once the client has been communicated to and a timeframe has been agreed.

Financial institutions may choose to "exit" or end a client relationship due to many reasons; anti-money laundering (AML), fraud, anti-bribery, corruption, sanctions exposure, reputational risk, non-profitability, and so on. To support this, financial institutions must have a suitable client selection and exit management framework.

3. WHY ARE CLIENT SELECTION AND EXIT MANAGEMENT IMPORTANT?

As financial institutions continue to grow, so does their client portfolio. With this increase, financial institutions must ensure relationships with clients are well managed to drive revenue and growth. Clients that do not generate enough business to cover their respective overheads impact the organization's bottom line. Financial institutions must consider how these unprofitable clients should be managed.

From a commercial perspective, establishing a solid client selection and exit management framework allows financial institutions to:

- Focus on expanding and building stronger relationships, product offerings, and commercial value with their profitable clients, by reducing the proportion of non-profitable clients.
- Reduce the cost of ongoing maintenance, remediation, and renewal exercises required to meet organizational and regulatory standards.
- Prioritize new clients and businesses that are better aligned to their risk range or profitability criteria. This is of paramount importance due to the ongoing cost of onboarding. A study conducted by Forrester Consulting, which measures the time, costs, and challenges involved in on-boarding institutional clients estimates that on-boarding can cost as much as U.S.$25,000 per client, with the average cost calculated at U.S.$6,000.[1] While U.S.$6,000 per new institutional client may not appear to be significant, when applied across different regions and locations

[1] http://bit.ly/2wgseuM

Table 1: Fines paid by major financial institutions in 2017.

FINANCIAL INSTITUTION	PENALTY	2017	REASONS
Credit Suisse	SGD$0.7 mln	May	The Monitory Authority of Singapore (MAS) imposed financial penalties for breaches of MAS Notice 626 – Prevention of Money Laundering and Countering the Financing of Terrorism.
United Overseas Bank	SGD$0.9 mln	May	
Citi Group	U.S.$97 mln	May	Willfully failing to file SARs
Bank of Ireland	€3.15 mln	May	Persistent breaches of Irish anti-money laundering and combating the financing of terrorism regulations.
Coutts & Co (Hong Kong branch)	HK$7mln	Apr	Breaching AML and terror finance laws.
Allied Irish Banks	€2.3 mln	Apr	Suspicious activity reports (SARs) and client due diligence (CDD) failings
Deutsche Bank	£163 mln	Jan	Serious AML failings
Deutsche Bank	$425 mln	Jan	Serious AML failings

it can become a very large figure indeed.

As Renato Ndokaj stated in his three-part blog on financial crime,[2] increased regulation aimed at tackling financial crime, including, but not limited to, AML, fraud, anti-bribery, and corruption, means that financial institutions must have greater visibility, ownership, and accountability of the risks held within their portfolio.

Financial institutions are continually adapting to these convoluted banking regulations, while at the same time trying to meet commercial pressures. A survey by Thomson Reuters, focusing on the cost and complexity of implementing Know Your Customer (KYC), found that financial institutions spend an average of U.S.$60 million to meet their KYC and Customer Due Diligence (CDD) compliance requirements, with some spending as much as U.S.$500 million.[3]

Institutions that can establish a strong client selection and exit management framework can reduce the costs, which can become quite steep, associated with clients that present financial crime risk or reputation risk to the firm. For example, In January 2017, Deutsche Bank was fined £163 GBP million by the U.K.'s Financial Conduct Authority (FCA) and U.S.$425 million by U.S. regulators over alleged trading schemes with Russia that operated from 2011 until early 2015 and resulted in laundering U.S.$10 billion out of Russia.[4] Whilst this fine was the largest financial penalty the FCA, and its predecessor FSA, has handed out for AML controls failings, Deutsche Bank are not alone. Financial institutions across the globe paid fines ranging from U.S.$500k to U.S.$500 million in 2017. Table 1 presents some of the fines paid by major global financial institutions.

4. THE GOAL

Client selection and exit management will vary from institution to institution, however, the end goal will typically be the same. Financial institutions should seek to develop relationships with clients that are low in cost and high in returns. Those who are deemed too costly (from both a monetary or non-monetary perspective) or have too low returns to retain the relationship should be considered for exit. Figure 1 presents the client management matrix that financial institutions need to consider.

Clients that are both low in cost and low in return and those who are high cost and high return would need further discussion and consideration before an exit decision should be made. Factors to consider include:

- Is further business expected in future?
- Is the client aware of the account (for accounts with

[2] http://bit.ly/2uTmqbc; http://bit.ly/2w0CjwJ; http://bit.ly/2tUjXfe
[3] http://tmsnrt.rs/2qm0thC
[4] http://on.ft.com/2jPJ65U

no transactions or revenue)?

- What would the impact of exiting the client in one location have with the overall relationship footprint?

Firms should articulate why the relationship should be retained and determine whether the benefits outweigh the costs.

Depending on the size of the organization, financial institutions need to decide whether to use a systematic or a manual process to determine whether a given client should be exited.

A systematic process that makes decisions based purely on commercial value may result in a near term increase in efficiency and cost savings, but it may also lead to reputational damage, which in turn could translate into long-term revenue losses. Given today's highly interconnected world, with online reviews as important in the eyes of many as the views of experts, an avalanche of complaints from unhappy long-term clients is not in the best interest of the organization.

In contrast, a highly manual and thorough review process for each exit case may put additional resource and cost constraints on the organization, and may offset any expected financial gains.

Figure 1: The client management matrix

		COST	
		LOW	HIGH
RETURN	LOW	Decision to be made	Exit
	HIGH	Retain	Decision to be made

Costs include monetary and non-monetary

Consequently, to terminate the accounts of non-profitable clients, financial institutions need to institute a systematic process that considers multiple factors, and not base its decision solely on financial inputs. They also need to ensure that there are sufficient controls and governance in place to allow for appeals and exceptions.

Due to its sensitive nature, potential exit cases related to financial crime or risk appetite should be put through a thorough review and discussion process. The risk of

falsely exiting a client relationship due to suspected financial crime is too high and could outweigh the benefits.

5. STAGES OF CLIENT SELECTION AND EXIT MANAGEMENT

To confirm clients are assessed, selected, and fully exited within a prescribed timeframe or framework, financial institutions must ensure they have sufficient controls and frameworks in place. At a minimum, a framework must be established and implemented to allow for the various stages or phases of an exit.

Table 2 outlines the fundamental stages required for client exits and the key considerations organizations must consider.

6. ESTABLISHING A SUITABLE FRAMEWORK

Both external and internal variables have a significant factor in determining the client selection and exit management strategy. There are, however, some fundamental steps that most financial institutions can or should adopt. This includes, but should not be limited to: goal and objectives, analysis and design, target and interim solution, implementation, and the transition to business as usual. Table 3 provides some consideration for each step.

Whilst frameworks can be used to address most cases, there will inevitably be scenarios that would not be covered. One example of this could be dividend accounts. Could a client account whose balance relates to dividend issued to shareholders be closed? The ultimate beneficiary owner to these funds are the underlying shareholders, not the client. Another example could be lack of client information, where no address or point of contact can be established, or they are out of date, but funds remain on the account. How and where would these funds be remitted to? Where would the relationship termination be sent to? When such events occur, these would need to be dealt with on a case by case basis. Appropriate follow ups and resolutions should be documented, as should the procedures.

Table 2: Fundamental stages required for client exits

STAGE OR PHASE OF EXIT	KEY CONSIDERATIONS	CONSIDERATION ACROSS ALL STAGES
Trigger/notification of potential client exit cases	• Identification of trigger sources • Systematic versus manual? • Do all stakeholders know how to notify potential client exits and to whom?	• How is information on terms of reference, procedures, policies, and framework shared or made available? • Who is notified or involved and at what phase and why? These include regulatory, compliance, relationship managers, financial crime, and banking or business heads. • Are there clear roles and responsibilities across different business areas and teams? • What is the governance and oversite model? Does something exist in each country of operation? • What is the exceptions management process? • Is there an end-to-end service level agreement across all stakeholders?
Building the information or case for discussion and decision-making	• Where are information sourced from? • How is information identified and shared across different business areas, such as retail banking, banking, securities? • Information assurance process?	
Decision-making and governance	• Approach – systematically or manual? • Governance – frequency and attendees? • How are client exits categorized? • Who's responsible for deciding and what's the escalation process when there is conflict?	
Execution on confirmed exit cases and reporting management information (MI)	• How does communication work (internal and external)? • Exit execution assurance process? • Escalation framework for breach in exit? • What level of MI can be produced?	

Table 3: Important steps for client exiting

OBJECTIVE/GOAL	ANALYSIS/DESIGN	SOLUTION/FRAMEWORK	IMPLEMENTATION/ EMBEDDING	BUSINESS AS USUAL
Scope • Retail banking/ institutional clients? Both? • Country, regional, or global? **Approach** • Regional teams? • Commitment from? • Governance committee? **KPI and success criteria** • What does success look like? • Increased profitability from clients? • Reduced operational costs?	**Organization structure and operating model** • Client footprint? • How are clients supported? **Current practices** • Process documents versus actual process adopted? • Technologies supporting process? • Communication internally and externally? **Internal & external risks/ issues** • Regulatory issues? • Regional or country issues?	**Target state, interim state, and minimum viable solution** • Technology to be developed and those ready to be used? • Process updates and gaps exiting in country or region? • Resources and support • Exceptions framework and issue management	**Stakeholder engagement** • Which stakeholders to include at which phase of the engagement? **Gaps current state versus interim/target state** • What model will be adopted? **Training and sign off** • Which medium to use for training? **"Go live" approach** • Phased? • "Big bang"? **Communication** • How will process, policy, and procedure updates be communicated internally?	**Governance and oversight** • Frequency? • Attendees? • Terms of reference? **Testing and continuity management** • Test scope? • Frequency? **Escalation and exceptions** • Methodology and template? • Approval? **Continuation and development** • Change manage process?

Table 4: Potential internal challenges

THEMES	OVERVIEW
Legacy systems	Through mergers and acquisitions, over time, the systems acquired and developed by financial institutions have become so complex that maintaining them presents significant operational costs and upgrading these systems would require a substantial amount of investment that would impact a firm's entire operating model. A study from NTT DATA Consulting, for example, revealed that financial institutions spend an average of 75%– 80% of their systems budget on maintaining legacy core deposit banking systems.
Poor data quality	Constantly changing requirements from regulators, business models, and product and service offerings by financial institutions has led to further enhancements and data attributes adding to the already complex system architecture embedded in organizations. As such, there is still a heavy dependency on people within organizations to validate data. For example, without the suitable data repository or source, identification of a unique client across multiple locations and business lines can be a challenging task. Is client "ABC123 Limited" in Hong Kong the same as client "ABC123 Ltd"? Exiting the relationship with the wrong client will impact the client experience and, even more concerning, may hinder the settlement of trades and transactions. Any financial institution making such mistakes may face reputational damage as well as financial penalties.
Obligations and contractual agreements	Long-dated transactions that cannot be novated, for example, also presents problems for the exits process. While other accounts and services may be blocked or closed, unless the trade can be novated the financial institution must meet the agreement and will only be able to close the account once the trade matures.

7. KEY CHALLENGES

While a minimum standard can be prescribed at a high level, most organizations will face unique challenges relevant to them. Though some organizations may have vigorous controls and may have some competitive advantages, the complex structure and nature of big financial institutions today means that there are inherent issues that may hinder the design, implementation, and running of a robust client selection and exit management framework.

Table 4 provides an overview of some of the internal challenges faced by organizations.

In addition to internal organizational challenges, financial institution also faces external challenges, such as local banking regulations that may impact their client selection and exit management strategy. In Malaysia, for example, clause 133 (1) of the Financial Service Act (FSA) 2013, stipulates that "133. (1) No person who has access to any document or information relating to the affairs or account of any customer of a financial institution, including the financial institution or any person who is or has been a director, officer, or agent of the financial institution, shall disclose to another person any document or information relating to the affairs or account of any customer of the financial institution"[5].

Depending on the scope or framework, this may present difficulty for organizations when implementing a global client selection and exit management model. If information cannot be shared on the client due to local regulations, would AML and sanctions be detected? If local exit decision was made, how would this be communicated and what impact would this have on the client relationship with other countries or regions?

8. COMMUNICATION

Prior to, during, and post exit decision communication is fundamental to any client selection and exit management strategy. This is the most important part of the process, since it ensures that the all the people who need to be informed internally are, and the message is clear and uniform across all departments when communicating with the clients whose accounts have been, or is about to be, terminated.

Internal communication, whether it is the relationship manager providing the client information to support or reject an exit decision, or the back office operation team closing the accounts, is crucial to ensure exits are appropriately managed. Internal communication should include, but is not limited to, KYC, AML, relationship managers, product owners, compliance, legal, and operations.

External communication is equally, if not more important than internal communication. Clients that have been confirmed as exit or to be exited, should be notified in an appropriate manner with clear explanation and rationale provided. As part of the discussion or notification clients must be given sufficient timeframe to make alternative arrangements. This may be through the form of a meeting, informal conversation, email, or a combination of the three. Formal communication and follow up, through a letter for example, should also be provided. These may be required for audit purposes, for example, and in such cases can be used by both parties.

9. CONCLUSION

Financial crime continues to be a key focal point for regulators and financial institutions alike. In addition to this, non-profitable clients continue to increase as organizations grow over time, adding unnecessary costs. Financial institutions need to provide sufficient investment into their client selection and exit management strategy to tackle these challenges.

Careful deliberation must be undertaken with regards to client selection and exit decision prior to the exit management. This article outlines the goal for organizations, key considerations, and the approach to client selection. However, financial institutions must choose an approach and framework that is fit for purpose for their organizations. Instituting an effective process and framework would enable financial institutions to better understand how new, or existing clients, align with their organization's profitability criteria and risk appetite.

As for the client exit management process itself, even with a developed strategy or framework, improved data quality, technological enhancements, and updated processes, human interaction is still a key dependency. To support any framework, all stakeholders must be engaged and understand their roles and responsibilities across the lifecycle of the client engagement. Communication plays a huge role in increasing or jeopardizing the success of a firm's client exit management approach.

The client selection and exit management landscape is still an evolving and challenging area. Failing to embed an efficient and robust strategy can negatively impact an organization (both monetarily and reputationally). Getting this right early on, however, will be very valuable. Financial institutions that can navigate through this landscape and gain competitive advantage will be able to reduce their exposure to high risk clients, mitigate their book of work (and therefore operational costs), and shift their focus to building better relationships with clients who better fit the organization's profitability criteria and risk appetite range.

[5] http://bit.ly/2w4XSfy

INVESTMENTS

PREVIOUS EDITIONS OF THE CAPCO JOURNAL OF FINANCIAL TRANSFORMATION ARE AVAILABLE AT **WWW.CAPCO.COM/INSTITUTE**

Intelligent financial planning for life

The hybrid advice model

Tax cuts: fuel share prices, not necessarily a catalyst for economic growth

Actively managed versus passive mutual funds: A race of two portfolios

Aligning interests over the long term: An incentive structure for U.S. 501(c)(3) private foundations

Financial inclusion and consumer payment choice

Intelligent financial planning for life

MICHAEL A. H. DEMPSTER | Professor Emeritus, Centre for Financial Research, Statistical Laboratory, University of Cambridge, and Managing Director, Cambridge Systems Associates[1]

ABSTRACT

This paper is concerned with the use of currently available technology to provide individuals, financial advisors, and pension fund financial planners with detailed prospective financial plans tailored to an individual's financial goals and obligations. By taking account of all prospective cash flows of an individual, including servicing current liabilities, and simultaneously optimizing prospective spending, saving, asset allocation, tax, insurance, etc., using dynamic stochastic optimization, this paper addresses the title by comparing the results of such a goal-based fully dynamic strategy with representative current best practices of the financial advisory industry. These include piecemeal fixed allocation portfolios for specific goals, target-date retirement funds, and fixed real income post-retirement financial products, all using Markowitz mean-variance optimization based on the very general goal of minimizing portfolio volatility for a specific portfolio expected return over a finite horizon. Making use of the same data and market calibrated Monte Carlo stochastic simulation for all the alternative portfolio strategies, we find that flexibility turns out to be of key importance to individuals for both portfolio and spending decisions. The performance of the adaptive dynamic goal-based portfolio strategy is found to be far superior to all the industry's Markowitz-based approaches. Superiority is measured here by the certainty equivalent increase in expected utility of individual lifetime consumption (gamma) and the extra initial capital required by an individual to put the dominated strategy on the same expected utility footing as the optimal dynamic strategy (initial capital gap). These empirical results should put paid to the commonly held view amongst finance professionals that the extra complexity of holistic dynamic stochastic models is not worth the marginal extra value obtained from their employment. We hope that such approaches implemented in currently available technologies will rapidly find acceptance by individuals, financial advisors, and pension funds to the genuine benefit of individual investors.

[1] This article is based on Dempster et al. (2016), whose authors I would like to thank for their painstaking and cheerful collaboration.

1. INTRODUCTION

Currently available technology can provide individuals, financial advisors, and pension fund financial planners with detailed prospective financial plans tailored to an individual's financial goals and obligations. By taking account of all prospective cash flows of an individual – including servicing current liabilities, and simultaneously optimizing prospective spending, saving, asset allocation, tax, insurance, etc., using dynamic stochastic optimization – the "iALM intelligent robo advisor" may be used to compare the results of a goals-based fully "dynamic" strategy with representative current best practices of the financial advisory industry. These include piecemeal fixed allocation portfolios for specific goals, target-date retirement funds, fixed real income post-retirement financial products, and commercial robo advisors, all of which use mean-variance optimization to address the very general goal of minimizing portfolio volatility for a specific portfolio expected return over a finite horizon.

iALM's design involves a synthesis of ideas and concepts. An objective data-driven, goal-based utility function is constructed from the personal data entered into the system by a user. Overall, the system's design is very much based on the personal situations, preferences, data, and flexibility that are the hallmarks of behavioral finance. The focus on personal finance comes from economics; models from mathematical finance are calibrated to historical data and used for asset return forecasting; network flow models from decision sciences are used to trace all of the model's cash flows; and stochastic optimization techniques are applied to solve a multi-level scenario-based problem under uncertainty over a very long horizon. Using patented software, the optimizer divides the planning period to the horizon into major portfolio rebalance points corresponding to the timing of a client's major financial decisions, such as house purchase and retirement. Even if you do not have control of the market, through iALM you can still have control of your life.

Making use of the same annual data and market calibrated Monte Carlo stochastic simulation for alternative portfolio strategies, in our experiments, as in our general experience, we have found that "flexibility" is of key importance to individuals for both portfolio and spending decisions. We have seen that the performance of the adaptive dynamic goals-based portfolio strategy is far superior to the industry's Markowitz-based approaches, as measured by the extra initial capital required by an individual to put the dominated strategy on the same expected utility footing as the optimal dynamic strategy. These empirical results should put paid to the commonly held view amongst finance professionals that the extra complexity of holistic dynamic stochastic models is not worth the marginal extra value obtained from their employment.

2. FINANCIAL PLANNING CHALLENGES

Financial planning for the benefit of individuals is based on a variety of approaches internationally. These range from simple heuristic approaches for selecting portfolios to approaches incorporating the joint stochastic optimization of asset allocation, contributions to different savings vehicles, and setting flexible saving and withdrawal rates. As affordable computing power and bandwidth continues to increase and the solution efficiency of large stochastic optimization problems expands, ever more complex financial planning tools are emerging. As we enter the age of big data, this trend will surely continue. Despite the relentless march of development, simple heuristic methodologies and mathematical approaches long criticized in the research literature continue to enjoy widespread acceptance by the financial planning industry. An important contributing factor to this divergence of approach is the difficulty of measuring and understanding the incremental benefits of incorporating more of the real-world complexities of household lifecycle financial planning. The results of complex stochastic modeling have gained most widespread acceptance by the general public in areas such as meteorological modeling, where it is easy for the man on the street to judge efficacy and benefit.

Some progress has nevertheless been made in measuring the benefit of different approaches to individual financial planning. For example, Morningstar introduced measuring the increase in the certainty equivalent income in moving from a benchmark approach to the approach recommended by an advisor, but this measure is arguably too abstract to gain popularity amongst clients purchasing financial advice. This article uses a new strategy comparison measure that is more intuitive, namely the "initial capital gap," which is the extra capital needed now to put the benchmark approach on the same expected utility footing as the recommended approach.

Here we analyze and decompose the value added by the stochastically optimized holistic goal-based

approach to financial planning embodied in the iALM "intelligent" robo advisor. For example, we measure the benefit of incorporating flexible "dynamic recourse" decision-making and test the strategy for individuals' saving toward retirement and other financial goals. The industry standard mean-variance approach to asset allocation incorporated in robo advisors, a fixed drawdown in terms of real income post retirement, and alternative savings vehicles are all considered. Our aim is to contribute to the understanding of whether the techniques currently used by the financial planning industry are inefficient in not making use of existing technologies and if so, how large these inefficiencies are.

A wide view is taken of what constitutes financial planning for individual benefit, including financial advisors meeting and advising individuals, the decisions of defined contribution trustee boards, and products marketed by the industry meant to address aspects of the lifecycle consumption problem, such as target-dated funds and living annuities. Significant differences exist in the best advice delivered for each focus and in the manner by which this advice is derived. In all cases, an entity claims to be an expert advisor, dispensing advice with a view to positively influencing the lifetime consumption of an individual household or many individual households. In a recent examination by the U.S. General Accounting Office of the common purpose in all these practices they were found wanting to varying degrees [GAO (2014)].

3. OUR STUDY

The literature on optimal investment strategies for retirement, and more generally optimal financial planning, is vast. However, the unique experiments discussed here aim to compare different solutions of the individual asset liability management problem within a common framework. For these experiments, two simple U.K. profiles were chosen: a young individual and a retired individual, both of whom are taken to be single for simplicity. We shall refer to these as Profiles A and B, respectively. The individual in Profile A is 30 years old, has no savings, earns £60k gross (equal to about £45k after tax) and has spending goals for "minimum," "acceptable," and "desirable" sterling amounts corresponding to 30k, 40k, and 50k, pre-retirement and to 7.5k, 40k, and 70k upon planned retirement at 65 (all in today's pounds sterling). The £7.5k per annum minimum amount post retirement represents the current U.K. subsistence level. The individual in Profile

B is 65 years old, has just retired and does not earn a salary. He has £600k in initial savings, and his post-retirement spending goals for minimum, acceptable, and desirable amounts correspond to 7.5k, 40k, and 70k.

We examine three types of solutions for these profiles:

- Solutions with various "static" asset allocations, "fixed" from the beginning, and only spending decisions being optimized.
- Solutions with "fixed spending" levels and only investment decisions being optimized.
- A fully "dynamic" solution with both investment and spending decisions being optimized.

Using multiple channels or portfolio wrappers with different tax treatments and asset allocation limits, portfolio addition and withdrawal (drawdown) amounts are set optimally in the fully dynamic solution. Our experiments decompose more granularly the value added by optimizing the optimal expected value of lifetime utility with this fully dynamic strategy and measure the benefits of incrementally incorporating:

- An optimal asset allocation informed by mean-variance optimization (MVO).
- Varying the level of risk of the mean-variance optimal strategy.
- Selecting an MVO strategy that is optimal with regards to a utility of lifetime income objective.
- Dynamic strategies that are only allowed to vary across time.
- Fully flexible dynamic recourse decision-making with path dependent decisions (allowing a different strategy depending on the Monte Carlo scenario up to the point of each decision.

Each of these steps adds complexity to the problem to be solved and all current solutions used in practice ignore one or more of these features to make the problem easier. The impact of each of these incremental complexities are very poorly understood by practitioners and they are often dismissed as unnecessary (or dubbed "spurious"), even by modeling experts. Such dismissals are not usually based on evidence, but they explain why the holistic features of iALM, or an equivalent approach, are deemed unnecessary.

4. iALM VERSUS MVO BASED ADVICE

The results presented here show the practical importance of these advanced features. In particular, we compare the fully dynamic iALM optimal strategy with the commonly recommended fixed MVO portfolio strategies, with and without fixed spending. Mis-specifying the optimal risk-return characteristics of the fixed MVO portfolio results in considerable losses to an investor's lifetime wealth. Tables 1 and 2 show that the detrimental effects of applying both fixed spending and fixed static portfolio strategies together is much worse than the sum of their individual fixed detrimental effects. For portfolios that are considered best from the perspective of expected utility over static MVO portfolios on the efficient frontier, that of the retired profile is quite close to the aggressive MVO portfolio, but for the young profile this optimal static portfolio is less aggressive than the corresponding aggressive MVO portfolio, showing that the "more risk" mantra is not always valid even when considering very long investment horizons. The "non-adaptive dynamic" solution adjusts portfolio asset proportions annually independent of the specific Monte Carlo scenario realizations and is a

Table 1: Initial capital gap to the dynamic solution for all strategies

STRATEGY	PROFILE A (000s)	PROFILE B (000s)
Non-adaptive dynamic	92	101
Static allocation – conservative	**1500**	600
Static allocation – moderate	350	280
Static allocation – aggressive	115	135
Fixed spending	18	200
Fully fixed	200	**1380**

Table 2: Certainty-equivalent lifetime spending for all strategies

STRATEGY	PROFILE A c-e SPENDING	PROFILE B c-e SPENDING	PROFILE A c-e SPENDING PER ANNUM	PROFILE B c-e SPENDING PER ANNUM
Dynamic	1,997,366	776,055	41,583	38,949
Non-adaptive dynamic	1,824,582	731,029	37,986	36,689
Static allocation – conservative	1,453,344	519,878	30,257	26,092
Static allocation – moderate	1,659,209	647,155	34,543	32,480
Static allocation – aggressive	1,818,123	721,629	37,851	36,217
Fixed spending	1,896,408	614,387	39,481	30,835
Fully fixed	1,724,720	413,477	35,907	20,752

Table 3: Gamma and gamma equivalent alpha of all strategies relative to the dynamic strategy

STRATEGY	PROFILE A GAMMA	PROFILE B GAMMA	PROFILE A GAMMA-EQUIVALENT ALPHA	PROFILE B GAMMA-EQUIVALENT ALPHA
Non-adaptive dynamic	9%	6%	0.21%	0.30%
Static allocation – conservative	37%	49%	0.73%	2.01%
Static allocation – moderate	20%	20%	0.43%	0.92%
Static allocation – aggressive	10%	8%	0.22%	0.37%
Fixed spending	5%	26%	0.12%	1.18%
Fully fixed	16%	88%	0.34%	3.14%

generalization of the life-staged fund strategy offered widely by the industry, in that its dynamic adjustments are made annually instead of periodically in life stages. In all cases, this dynamic strategy outperforms all the static asset allocation strategies. The dynamic iALM strategy, however, achieves even higher lifetime utility through dynamic management of all cash flows. For our experiments, a tolerable portfolio annual loss constraint of 15% was introduced and compared with a portfolio loss tolerance of 100%, i.e., the no portfolio loss penalty, which is used in standard risky advice. The lack of sensitivity to the portfolio loss tolerance of the iALM fully dynamic strategy suggests that with this optimal strategy there can be a cap to the risk of portfolio loss at no significant cost to expected lifetime spend.

Our overall findings are perhaps best understood by the results of the comparative value calculations relative to the dynamic strategy for all the alternative strategies we have evaluated. The strategy comparisons by our initial capital gap comparison measure are reported in Table 1, which shows, for example, that the young individual employing a conservative fixed MVO portfolio strategy would need an initial windfall of £1.5 million to expect to achieve with this strategy the same utility of lifetime consumption as the dynamic strategy would yield with no initial capital. Even for the aggressive static portfolio strategy the initial extra capital needed is seen to be significant for both profiles. Perhaps the worst situation revealed by this measure is that of the just retired individual following a fixed MVO aggressive portfolio strategy with a fixed post retirement drawdown. The retiree would need an extra 1.38

million pounds in savings at retirement to match the prospective expected results of the dynamic strategy for their accumulated £600,000 savings.

Table 2 contains the results of the certainty equivalent lifetime spending calculations necessary to compute the Morningstar "gamma" strategy comparison measures, and their per annum values. Focusing on these more easily interpreted annual spending values, we see that fixed spending for the young profile while earning a salary is not a massive burden. The results in Table 2 are in the expected order and do not differ markedly from the two risk tolerances, except for the fixed MVO aggressive strategy, which is consciously risky. The superiority of the full dynamic strategy over all others, including the non-adaptive dynamic strategy, is in clear agreement with the results in Table 1. The fully fixed strategy, although based on the MVO Aggressive portfolio strategy, is poor for both profiles and particularly bad for the retired Profile B. The strategy comparison results, in terms of gamma and "gamma-equivalent alpha" (the extra per annum portfolio return

Table 4: Probabilities of achieving the £40k acceptable target spending level

STRATEGY	PROFILE A PRE-RET	PROFILE A POST-RET	PROFILE B
Dynamic	45%	80%	70%
Non-adaptive dynamic	17%	54%	54%
Static allocation - conservative	5%	0%	0%
Static allocation - moderate	1%	25%	10%
Static allocation - aggressive	15%	50%	50%
Fixed spending	0%	45%	0%
Fully fixed	0%	55%	0%

Figure 1: Young profile dynamic strategy prospective expected asset allocation

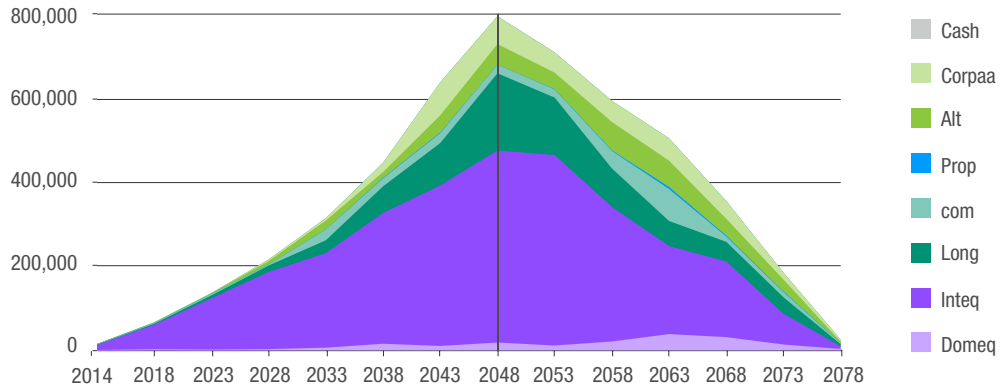
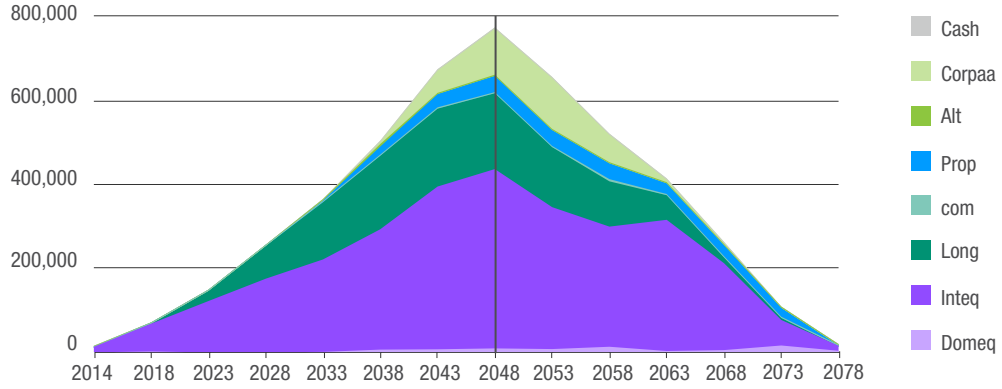


Figure 2: Young profile non-adaptive dynamic strategy prospective expected asset allocation



needed to match the certainty equivalent lifetime spend of the dynamic strategy) are given in Table 3.

Many practitioners will concede that liability optimized approaches make sense, but that such complex analysis is unnecessary because they believe that investors should employ a highly aggressive strategy that will, in the long run, deliver the best results, even when considered in terms of income. These results show that this is simply not true. For the young individual, the dynamic strategy outperformed the static aggressive strategy by well over £100,000 on an initial capital basis and by 10% on a gamma basis. Very few households would deem these differences negligible. The results also show just how detrimental strategies that are traditionally thought of as "conservative" can be.

Table 4 shows the achievement probability results for the target (acceptable) spending level of £40k per annum.

The success probabilities for the static moderate strategy are vastly worse than those for the static aggressive and dynamic strategies. The same holds true for the static aggressive strategy relative to the dynamic strategy. The success probabilities shown in Table 4 represent a powerful and tangible demonstration of just how much difference advanced planning techniques can make. Many modeling practitioners in industry realize that the fully dynamic approach to financial planning is an improvement on current practice and is actually implementable today, but they often jump to the conclusion that its value added will be very marginal. These results for the young and retired profiles show indisputably the value added by fully dynamic strategies. There is no rational debate to be had regarding changes of target achievement from 15% to 45%, or 50% to 70 or 80%, not being significant.

Table 5: Contrast of intelligent versus standard robo-advice

	STANDARD ROBO-ADVISOR	iALM INTELLIGENT ROBO-ADVISOR
Holistic optimization: • Goals (children's education, mortgage, etc.) • Taxes, transaction costs, fees, etc. • Every **goal influences decisions** on all other goals	✗	✓
Future **dynamic** portfolio allocation	✗	✓
Advice on **how much to save**	✗	✓
Accounts for **longevity risk**	✗	✓

Finally, by way of illustration of the iALM intelligent robo-advisor's screens we present for the young Profile A the prospective expected portfolio evolutions over the lifecycle from the initial portfolio allocations. Figures 1 and 2 show these prospective future expected portfolio evolutions corresponding to the dynamic and non-adaptive dynamic portfolio strategies taking account of all transactions costs. Retirement dates are shown by the vertical lines

The overall shape and quantities of the prospective asset allocations over the lifecycle differ quite significantly. There is a far larger allocation to long bonds in the non-adaptive dynamic strategy, because the dynamic strategy has far more de-risking/hedging flexibility. In Figure 2, the prospective expected allocations for the young non-adaptive dynamic strategy look very similar to the heuristic rule of gradually decreasing the share of equity and increasing the share of bonds in the portfolio over an individual's lifetime. The non-adaptive dynamic framework thus generates a life-staging approach to prospective portfolio evolution, but is less effective than the fully dynamic approach for which the equity to bond shift pattern in Figure 1 is significantly less prominent.

Put simply, non-dynamic strategies are not realistic representations of how people actually approach the lifetime consumption problem, as they ignore the interventions that investors will undoubtedly make. Table 5 contrasts the main features of current fixed standard robo-advice with the dynamic intelligent robo-advice of iALM.

5. CONCLUSION

In summary, the iALM intelligent robo-advisor has been employed to demonstrate the positive effects of using a dynamic stochastic goal-based holistic approach to address the lifecycle consumption problem. We decomposed the relative value-added for individual clients or pension fund members using this technology to evaluate the current bases of advice given by the advisory industry to clients. Such advice includes Markowitz mean-variance optimized portfolios with varying degrees of risk aversion; specific goal funds, for example, to cover an individual household's future school or university fees; life-staged funds; and fixed real post-retirement spending, by means of fixed defined contribution pension fund withdrawals or the purchase of an indexed fixed annuity at retirement.

Each of these industry standard bases was embedded in the iALM dynamic stochastic planning system and their relative effectiveness in meeting an individual's goals was evaluated by means of two comparative statistics. Both statistics, "initial capital gap" and "gamma," were based on the optimal expected utility of lifetime consumption and supplemented by spending target achievement probabilities and prospective future portfolio evolutions. The results are surprising, even to us, as the dynamic flexibility embodied in the holistic iALM model significantly outperforms the other approaches – fixed post-retirement spending in real terms being particularly bad. These results will hopefully go some way to convincing the pensions and financial advisory industry and regulators that it is worth the extra effort to employ the dynamic holistic stochastic strategies required to address members and clients' actual needs.

References

Blanchett, D., and P. Kaplan, 2013, “Alpha, beta, and now… gamma,” Journal of Retirement 2:1, 29-45

Dempster, M. A. H., and E. A. Medova, 2011, “Asset liability management for individual households (with discussion),” British Actuarial Journal 16, 405-464

Dempster, M. A. H., and E. A. Medova, 2012, “Planning for retirement: asset liability management for individuals.” in Mitra, G., and K. Schwaiger (eds.), Asset liability management 2012 yearbook, Palgrave Macmillan, 409-432

Dempster, M. A. H., D. Kloppers, E. A. Medova, I. Osmolovskiy, and P. Ustinov, 2016, “Life cycle goal achievement or portfolio volatility reduction?” Journal of Portfolio Management 42:2, 99-111

GAO, 2014, “401(K) Plans: improvements can be made to better protect participants in managed accounts,” U.S. Government Accountability Office, June, Available at http://bit.ly/2xH0zXu

Medova, E. A., J. K. Murphy, A. P. Owen, and K. Rehman, 2009, “Individual asset liability management,” Quantitative Finance, 8:6, 547-560

Moore, H., M. A. H. Dempster, D. Kloppers, E. A. Medova, I. Osmolovskiy, P. Ustinov, 2016, “Practical applications of life cycle goal achievement or portfolio volatility reduction?” Practical Applications 4:1, 43-46

The hybrid advice model

KAPIN VORA | Partner, Capco Digital
TOBIAS HENRY | Managing Principal, Capco Digital
JACOB WAMPFLER | Senior Consultant, Capco
MIKE CLARKE | Senior Consultant, Capco

ABSTRACT

Wealth management has traditionally meant a dedicated human financial advisor, providing tailored investment advice to primarily high-net-worth and ultra-high-net-worth clients. Over the past 10 years, innovations in financial technology, increased regulation, and changing generational expectations have challenged our understanding of the client-advisor relationship, and the wealth management industry as a whole. The rapid growth and proliferation of financial services technology firms has caused a shift in the market. The digital first, low-cost alternatives they provide have opened wealth management to a broader base of retail, affluent, and mass affluent customers.

While the wealth management debate has focused on digital advisor versus human advisor, a third option has emerged, the hybrid advisor model, which combines the best aspects of both traditional human advice and automated digital platforms. Firms that successfully employ the hybrid advisor model will be empowered to evolve their business, attract a new generation of customers, and serve them in a flexible and cost-effective manner, while realizing greater agility in their business models, technology, and product offerings.

1. INTRODUCTION

Wealth management has traditionally referred to working with a dedicated human financial advisor, one who provides tailored investment advice to primarily high-net-worth and ultra-high-net-worth clients.

However, over the past 10 years, innovations in financial technology, increased regulation, and changing generational expectations have not only challenged our understanding of client-advisor relationships, but altered our perceptions of the wealth management industry as a whole. In this relatively short period of time, we have seen exchange traded funds (ETFs) replace mutual funds, digital channels overtake traditional interaction methods, and a booming global fintech industry challenge incumbents on providing better products and services to the same customer base.

The rapid growth and proliferation of financial services technology firms has caused a shift in the market by providing digital-first, low-cost alternatives. This has opened wealth management to a broader base of retail, affluent, and mass affluent customers.

While the wealth management debate has traditionally focused on digital advisors versus human advisors, a third option has recently emerged: the hybrid advisor model. This new model interweaves the best aspects of both traditional human advice and automated digital platforms into one seamless experience.

2. WEALTH MANAGEMENT AT A TIPPING POINT

In 2008, a myriad of financial services technology companies entered the market, each launching their own digital-only, direct-to-consumer robo-advice solutions. These companies mainly offer low-cost ETF and mutual fund-based investment solutions with no-to-low investment minimums. Their mass affluent customer base, combined with their seamless adoption into the marketplace, has highlighted that wealth management is not just for the wealthy.

These firms have also pioneered a digital-first model within wealth management, one focused on ensuring the efficiency and ease of user experience. The success of these robo-advisors has proved to traditional wealth managers that it is possible to alleviate current customer pain points through digitization. Digitizing areas such as onboarding, document transfer, portfolio analysis, and performance tracking have all proven to improve customer experience and reduce operating costs.

There is no question that these new entrants have paved the way in defining the future of wealth management. As their clients and assets under management (AUM) have expanded, leading digital firms have even been seen as genuine competition for their traditional wealth management counterparts. However, the current data suggests many of these firms have saturated their market, experiencing reduced AUM growth over the past 24 months.

> "The ability to provide tailored human guidance, along with cutting edge digital expertise, can be a key differentiator in gaining the upper hand."

One potential explanation involves consumer comfort with tried-and-true methods of wealth management. For instance, our research suggests that customers are not comfortable with investing large sums of money (more than U.S.$100,000) in a solution that is digital-only. When it comes to their wealth, customers seem to want some level of human interaction to help guide them through complex financial decisions and product options facilitated by an exclusively-digital experience. Consumers also want flexibility in the way they interact with their financial institution. Ergo, a one-size-fits-all model is a thing of the past – even in the high-tech world of digital.

Many digital advice firms recognized this trend, and have quickly pivoted their business models to focus on enabling financial advisors to connect with consumers through their platforms. They realized that there was more potential for growth if they partnered with large financial institutions that offered pre-existing customer bases with a larger pool of assets. In return, the financial institutions would receive a white-labeled version of a digital platform, one tailored to their business that could help accelerate their digital agenda.

Furthermore, established robo-advisor firms have chosen to pivot by supplementing their existing digital-only offerings with human advisors. Firms have found that customers still want access to a human advisor for guidance and advice, especially as their financial

needs become more complex. Customers also want their advisor available as needed through the channel of their choosing. However, these services come at a cost to the customer, with increased fees and higher minimum investment requirements.

Large financial institutions recognize that their traditional business model, which has gone unchanged for over 100 years, is being challenged in a major way. The fees they charge, channels they engage their customers through, and client segments they wish to attract, are all shifting day by day. In the face of this turbulence, firms need to understand that choosing not to embrace digital creates a tangible risk of being left behind.

While traditional wealth management firms have been slow to adapt, many are gradually seeing the value of using the technology provided by digital advice firms. The ability to provide tailored human guidance, along with their cutting-edge digital expertise, can be a key differentiator in gaining the upper hand.

In early 2015, a surge of acquisitions and partnerships began as wealth managers sought the best digital advice platforms for modernizing their businesses. The new medium they sought would automate core functions of investment and account management, and enable the advisor to focus on both holistic financial planning, and building relationships with clients.

Whichever avenue they take, one thing is common: firms of all sizes see the hybrid approach as a pillar of their growth strategy, something that can ensure their status as top-tier wealth managers for this generation and the next.

3. THE HYBRID ADVICE MODEL

In its simplest form, the hybrid approach combines the best components of human-based financial advice and digital advice, offering a flexible and tailored wealth management solution to clients of all demographics.

The hybrid advice solution is underpinned by a flexible business model that can support customers throughout their financial lives, spanning from mass market to ultra-high-net-worth. The hybrid approach has three models to offer to customers, which depend on their customer segment (defined by investable assets), as well as the complexity of their financial needs.

The level of human interaction, product complexity, fees, and accounts offered changes between business models. The business models shown in Figure 1 indicate the optionality of the solution, and illustrate the flexibilty of the hybrid approach to meet the needs of all consumer segments.

Figure 1: Hybrid advisor model

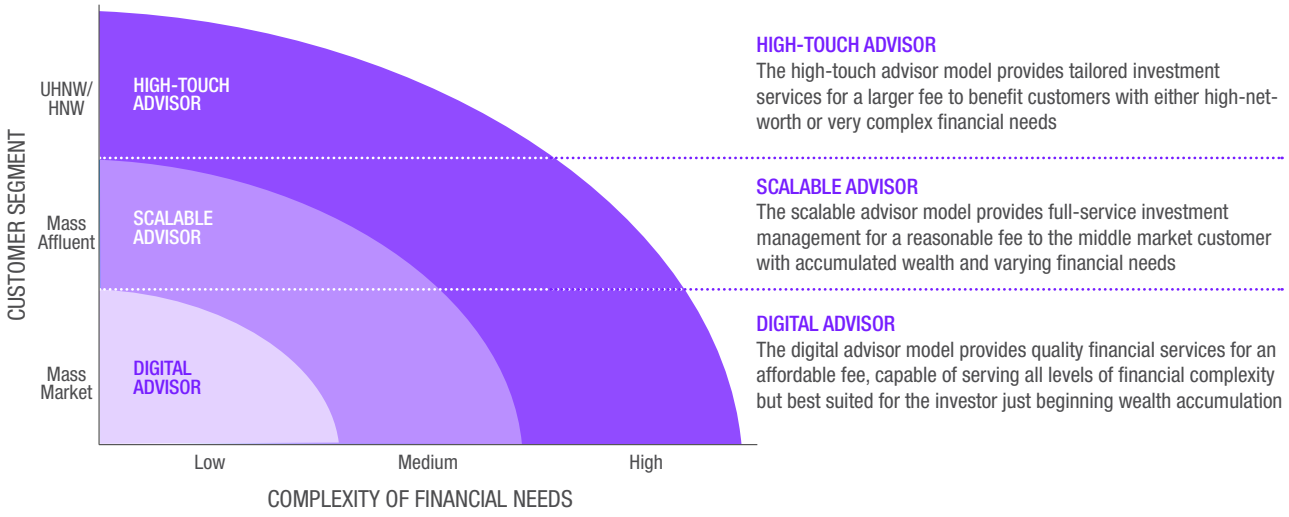


Table 1: Annual proportion of interest income resulting from term transformation

	DIGITAL ADVISOR	SCALABLE ADVISOR	HIGH-TOUCH ADVISOR
	Provides mass market investors access to affordable, quality financial services	Enables advisors to offer mass affluent investors financial services suited to their needs	Offers investors with complex financial needs a dedicated and tailored advisory experience
ADVISOR ROLE	Ongoing access to call center support model	Advisor guidance during account opening, Ongoing access to advisor guidance as needed	Dedicated human advisor guidance from account opening to ongoing portfolio management and reporting
DIGITAL FEATURES	On-boarding, omni-channel access, account funding, model generation, investment management, rebalancing, tax-loss harvesting, and reporting	On-boarding, omni-channel access, account funding, model generation, investment management, rebalancing, tax-loss harvesting, and reporting	On-boarding, omni-channel access, account funding, reporting
ACCOUNT MINIMUM	U.S.$0 – U.S.$100,000	U.S.$100,000 – U.S.$500,000	U.S.$500,000+
COST TO CUSTOMER	Low (e.g., 0-30bps)	Medium (e.g., 30-150 bps)	High (e.g. 150bps+)
ACCOUNT TYPES	IRAs, retirement, goal-saving	IRAs, retirement, goal-saving, financial planning, 529s	IRAs, retirement, goal-saving, financial planning, 529s, estate planning, insurance, lending, healthcare
PRODUCT OFFERING	ETFs, equities	ETFs, equities, mutual funds, fixed income	ETFs, equities, mutual funds, fixed income, real estate, insurance, annuities, alternative investments

3.1 The hybrid advisor value chain

Managing investments using automated technology enables advisors to grow their existing businesses by focusing on financial planning and customer relationships, while scaling their business to serve a larger customer base. With flexible business models, advisors can engage customers at a younger age, and continue to provide cost-effective services to them as they move through different stages of life. Digital capabilities also dramatically enhance the advisor value proposition by improving communication through both mobile applications, and on-demand access to portfolio performance.

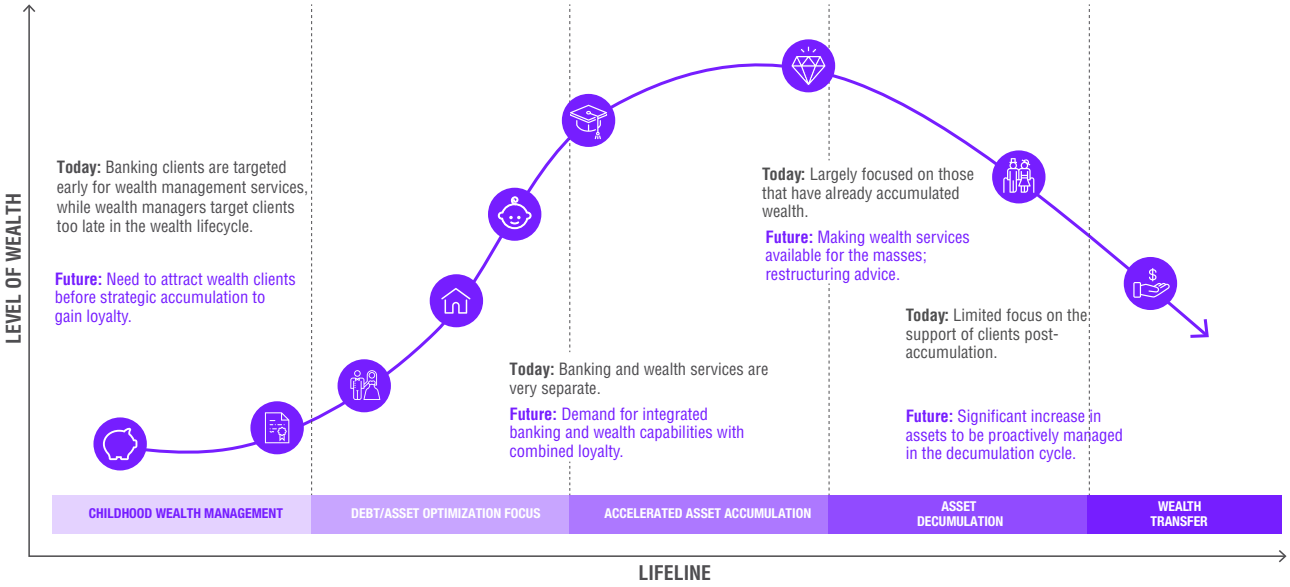
Today, people have become so accustomed to having information readily accessible to them and finance should be no different. The hybrid model facilitates this modern accessibility and ease. It allows clients to benefit by having instant access to real-time financial data through their desktop and mobile devices, while instant messaging and chatbots allow them to connect with advisors on demand. This increased connectivity has the added benefit of encouraging more frequent interactions between clients and advisors, which strengthens their relationship and mutual regard. Centralizing all financial information into one platform has finally allowed clients to plan and monitor portfolios – across all accounts, products, and investment solutions – at any time they please. Moreover, these upgrades in business model flexibility have also yielded flexibility in fee structures, thereby allowing clients to pay for the exact level of service they would like.

3.2 Variants of the hybrid advisor model

Multiple hybrid business models enable advisors to efficiently offer services based on customer segment, and complexity of financial need. The “digital advisor,” “scalable advisor,” and “high-touch advisor” variations allow advisors flexibility to serve clients across their entire wealth lifecycle.

For a mass market investor, value is driven from simplicity, and the ability to easily interact. The mass affluent do not derive value or make decisions based on performance – not because they don’t want to make money, but rather because they lack the financial expertise to do so. Northwestern Mutual reported that a third of Americans do not have a financial plan (2015). This suggests that the problem lies not in a shortage of tools, but rather in the ability to use them.

Figure 2: The wealth lifecycle



The digital advisor model targets mass market investors seeking affordability and quality financial services. Here, human advisors only provide initial guidance and setup, after which a digital platform takes over for on-boarding, omni-channel portfolio access, account funding, model generation, investment management, and rebalancing. Given their ability to easily serve a large customer base, products offered in this model are quite like those historically offered to the mass market, such as ETFs, equities, and cash management. However, investors with digital advisor accounts can seamlessly graduate to scalable or high touch service levels as assets and financial complexities grow.

The scalable advisor model allows both robo and traditional advisors to serve mass affluent clients in an efficient and cost-effective manner. With the scalable model, advisors provide a more active role in investment decisions, and help manage complex financial instruments in a client's portfolio. Investors, therefore, receive both an enhanced digital experience and full-scale advisor services.

Many say that once a client accumulates enough wealth to be considered ultra-high-net-worth (UHNW), there is no longer a role for digital in the wealth management experience. The reality is that over 60% of UHNW individuals are under 45 years of age. As the investor begins to grow assets and contend with more complex life events, even HNW and UHNW individuals will seek out the ability to get answers to complex questions anytime and anywhere – something that reaches beyond the scope of measured advisory visits.

A high-touch advisor model is specifically designed to provide investors with complex financial needs with a tailored advisory experience. Advisors not only encompass the traditional offerings of lower-tier services, but provide clients with personalized assistance on financial, retirement, and estate planning. Product scope for high-touch advisors is expanded to include real estate, annuities, and alternative investments. This allows advisors to both leverage their client's assets to hedge, and invest using more exotic solutions. Advisor-client interaction is also not solely limited to digital channels, and often requires a much higher frequency of in-person meetings to establish trust and demonstrate a dedicated focus on the client's financial picture.

The hybrid model has many inherent benefits that spans across all parties, which make it an attractive and mutually beneficial option for financial advisors, businesses, and clients alike.

The greatest benefit of the hybrid approach is that it increases the scalability of financial advisors, therefore allowing them to attract and serve more clients while maintaining high quality service. This is mainly because the implementation of a digital platform helps to automate manual and time-consuming processes for both client and advisor.

The most material client benefit of the hybrid approach lies in how it allows them to choose their level of human and digital interactions. The solution provides flexibility, allowing the client to choose their level of advice, product access, fee structure, and digital experience. Hybrid enables a personalized solution that is cost-efficient to the business and provides material growth opportunity for the financial advisor.

The digital platform also acts as a catalyst for driving business growth through acquisition of new customers and, subsequently, new assets. A digital offering provides a low-cost feeder channel to attract millennial customers, who have huge earning potential over time.

The hybrid model enables advisors to attract new assets from new and existing customers, and, equally important, allows the advisor to retain assets currently managed. This is because a fundamental principle of the hybrid approach flexible business model lies in its ability to allow advisors to service customers regardless of their demographic, available assets, or stage in financial life. As a client accumulates more wealth, and their financial needs become more complex, the hybrid model allows advisors to efficiently transition clients from a digital-only experience to one with more human interaction and enhanced services.

As advisors realize efficiencies of scale, they can also see how implementing a digital platform provides the parent business with enhanced transparency and control across the value chain. Supervisory regulation within financial services requires firms to ensure products and services meet the needs of their clients. Digital platforms provide automated guidelines to enhance monitoring, supervision, and risk scoring.

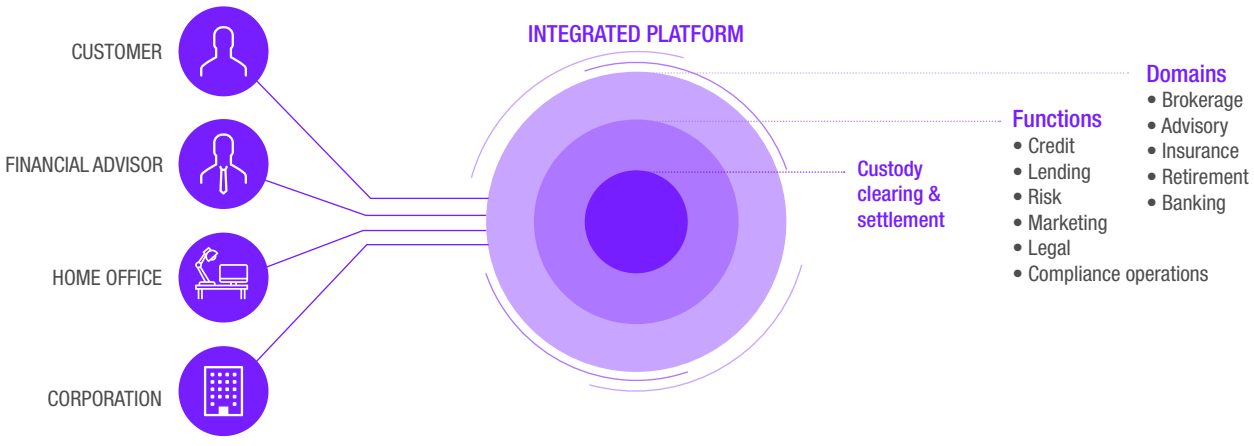
4. UNIFYING THE CUSTOMER EXPERIENCE

The hybrid approach has the potential to provide a unified platform accessible by the financial advisor and the client, thereby providing a consistent experience. In addition, the hybrid approach enables channel flexibility across desktop, mobile, and tablet. Furthermore, solutions that use cloud-based technologies allow customers to seamlessly move between platforms with no impact to the user experience. In the future, this platform unification will provide the client with a holistic view of their wealth across multiple providers and products.

The hybrid approach is enabled by a digital platform that can cater to all demographics, regardless of assets and complexity – one flexible enough to evolve at the rate of technology, industry, and customer demographic change.

The digital platform enables a truly multi-channel experience, with an open API architecture that allows

Figure 3: The integrated digital platform



GOALS-BASED PLANNING

Enables profiling and solutioning to be done based on a client's investment goals by removing product silos and providing a view into the client's full financial picture.

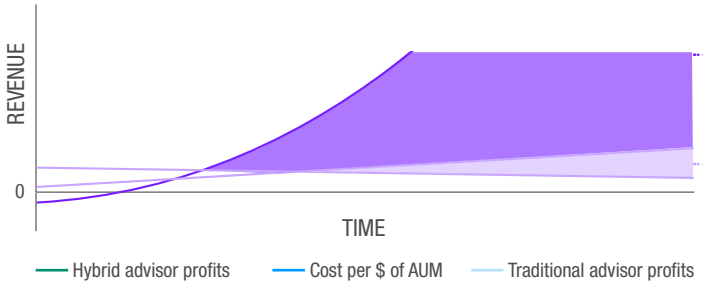
RELATIONSHIP FOCUS

Allows advisors to create holistic financial plans that span all of a client's financial needs, taking into account investable and held-away assets.

ANALYTICS AND RISK OVERSIGHT

Transforms the way data is collected, managed, analyzed, and employed within the firm providing a consistent approach to portfolio monitoring and supervision.

Figure 4: Hybrid advisory versus traditional advisory profit potential



PROFITS PER DOLLAR OF AUM (HYBRID):
Profit potential for the hybrid advisory model is massive due to increased advisor capacity, lower operating costs, and efficient servicing of clients.

PROFITS PER DOLLAR OF AUM (TRADITIONAL):
Traditional advisors are limited in scale and have experienced a loss in profitability as fees have been driven down by client expectations and regulations.

for seamless integration with the fintech ecosystem, as well as existing legacy infrastructures. The digital platform provides a single point of entry for all users, helps reduce process inefficiency, opens up new products, and improves supervision and control.

A single point of entry allows advisors, clients, home offices, and corporations to interact through a common gateway defined by a holistic and consistent experience. The digital platform should be built in a way that supports multiple asset classes, products, and account types. This unified view of a client's wealth allows advisors to spend less time monitoring portfolios on separate platforms, and more time cultivating client relationships and managing their wealth.

Perhaps most importantly, in an environment with heightened regulatory scrutiny, the digital platform can transform the way data is collected, managed, analyzed, and employed within a firm. This enables a consistent approach to portfolio pricing, monitoring, and supervision.

5. BUILDING THE BUSINESS CASE

There are three main avenues to explore prior to undertaking a digital transformation and transition to the hybrid advice approach: partner, build, or acquire. The adoption of each of these avenues will be influenced by the firm's culture, budget, and appetite to disrupt.

5.1 The case for acquisitions

Acquisitions can offer a variety of benefits, including speed to market, increased profitability, and exclusivity, as well as greater strategic opportunities. With a newly acquired digital platform, wealth managers can enter new markets faster than with a build option, and immediately convert new or existing customers to a lower cost hybrid model.

Buying established companies also allows complete control over any proprietary technology, hence allowing these firms to offer clients a completely unique and exclusive experience.

Despite a faster go to market timeline, acquisitions can be costly. They also demand a greater integration effort than partnerships with "out-of-box" solutions. Given the rate of robo-advisor venture capital funding, there also exists the risk of inflated valuations, which may force firms looking to acquire new technology to pay a premium.

5.2 The case for building

Building an in-house digital platform is an appropriate solution for firms with both the development capacity, and the dedicated strategic direction towards digital innovation. Vanguard, Merrill Lynch, Charles Schwab, and TD are all examples of firms that have done so, each having built their own proprietary digital platforms. Control is a critical pillar of the build strategy. Firms will have ownership of their own intellectual property, functionality, customer experience, and data. Furthermore, pricing, margins, and revenue potential will fall under complete command of wealth managers.

While a build strategy offers attractive control over the full IT stack, it also has the longest time to market of all possible scenarios. Stakeholder consensus, regulatory approval, and deep functional and technical knowledge can be difficult to obtain. Furthermore, given that there is no promise of immediate increased revenue, the initial cost can make it difficult to justify the price tag of building an in-house digital platform.

5.3 The case for partnering

Partnering is the fastest, cheapest, and least resource-intensive solution for constructing a digital wealth management solution. White-labeled platforms offer easy out-of-box integration, and often come pre-integrated with major custodian banks. Partnerships also allow for immediate customer acquisitions because they allow access to the partner platform's existing clients. For example, UBS and Wells Fargo have each embarked on partnerships to provide robo-advisor experiences to clients, while Raymond James and John Hancock each forged partnerships that allow them to offer both robo and hybrid advisor solutions to clients.

There are, however, lingering challenges with this approach. Partnerships limit a firm's control over system functionality, and can reduce the competitive advantage that a build or buy strategy offers. A partner strategy can also decrease revenue margins, as firms will be expected to pay licensing fees or share profits.

5.4 Economics

Although the choice to build, buy, or partner is highly dependent on each firm's differing long-term strategies, there are clear economic drivers for implementing a digital platform.

By optimizing the balance between revenue and cost drivers, hybrid models can help deliver a business strategy that increases profitability while decreasing operating costs. Through an expected increase in AUM, innovation acceleration, and an expanded customer base, the hybrid model offers multiple opportunities to drive revenue, while simultaneously diminishing operational and technology costs, reducing on-boarding costs, and lowering the overall cost of acquiring a customer (CAC).

5. THE FUTURE IS NOW FOR DIGITAL ADVICE

Independent of the form, digital advice and technology are the future of wealth management. Providing a digital experience has become a necessity for traditional wealth management firms. Customers and financial advisors expect better digital tools and the flexibility of human interaction to suit their specific needs.

Embracing the fintech ecosystem and using digital advice platforms to provide a hybrid approach, empowers wealth management firms to evolve their business. It can not only attract a new generation of customers, but serve them in a flexible and cost-effective way that will ensure the longevity of their relationships.

Although the future is bright for wealth managers who successfully complete this shift, pressure remains for continued innovation. Firms like Google, Facebook, Alibaba, and Apple, each already a leader in the digital marketplace, all pose massive threats to the financial services industry.

Additionally, the regulatory landscape since the 2008 financial crisis has encroached on traditional business models – as illustrated by the Dodd-Frank Act and the Department of Labor's Fiduciary Rule – while the SEC recently placed automated digital advice as a top examination priority for 2017 and 2018.

Firms that successfully employ the hybrid advisor model, one powered by a digital platform, will have the competitive advantage of greater agility in their business model, technology, and product offerings.

It is becoming clear that the changes happening in wealth management will benefit clients and advisors alike, and help democratize access to quality financial advice. Investment solutions once reserved for the ultra-wealthy will begin to trickle down and serve clients without distinction. Clients will be empowered to help drive their own financial independence, and advisors will grow their business by offering holistic financial advice, all of which culminates in the ideal wealth management experience.

Tax cuts: Fuel share prices, not necessarily a catalyst for economic growth[1]

BLU PUTNAM | Chief Economist, CME Group
ERIK NORLAND | Senior Economist, CME Group

ABSTRACT

Tax cuts are typically politically popular. And, they are often justified by their potential to stimulate economic activity; the concept being that lower tax rates lead to higher real GDP growth, and down the road higher tax revenues. While tax rate reductions seem to support equity prices, the link between lower tax rates and future economic growth is exceedingly tenuous. Economic theory sees lower marginal tax rates as driving more investment and economic activity; however, such an outcome depends on whether meaningful tax reform and simplification accompanies the marginal tax rate cuts. Unfortunately, meaningful tax simplification and reform rarely make it through the political process. Hence, loopholes, not tax cuts, continue to drive investment decisions, meaning that tax rate reductions often disappoint in terms of the political promise of higher future economic growth. Looking forward, if the U.S. goes ahead with large corporate and personal income tax cuts effective in 2018, we see little prospect for higher real GDP growth resulting from any tax reductions because we are pessimistic about tax simplification. We do see tax reductions adding materially to the U.S. debt load. Indeed, tax cuts leading to higher debt loads might cause the Federal Reserve to be overly cautious on raising rates, which could negatively impact the U.S. dollar.

[1]Disclaimer: All examples in this report are hypothetical interpretations of situations and are used for explanation purposes only. The views in this report reflect solely those of the authors and not necessarily those of CME Group or its affiliated institutions. This article and the information herein should not be considered investment advice or the results of actual market experience.

1. INTRODUCTION

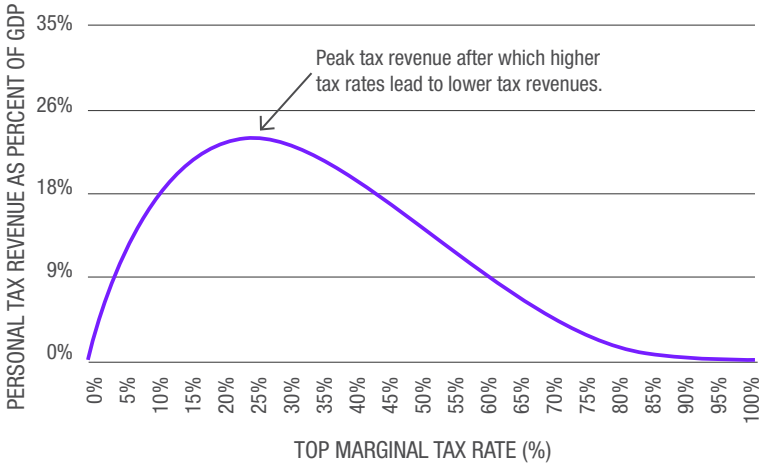
Tax cuts are typically politically popular. They are often justified by their potential to stimulate economic activity; the concept being that lower tax rates lead to higher real GDP growth, and down the road higher tax revenues. The link between lower tax rates and future economic growth is exceedingly tenuous. The statistical evidence for tax cuts leading to higher economic growth is mixed and not very convincing. A number of the critical assumptions in the economic theory of tax cuts are often ignored. When one replaces these heroic assumptions with a more realistic view of the world, it goes a very long way to help explain why tax cuts do not seem to contribute to economic growth, when the intuition is otherwise. Even so, and despite the lack of impact on economic growth, unambiguously, tax cuts seem to help raise share prices.

Our research focuses, firstly, on the economics of tax-rate reductions. Economic theory sees lower marginal tax rates as driving more investment and economic activity; however, such an outcome depends on whether meaningful tax reform and simplification accompanies the marginal tax cuts. Unfortunately, meaningful tax simplification and reform rarely make it through the political process. Hence, loopholes, not tax cuts, continue to drive investment decisions, meaning that tax rate reductions often disappoint in terms of the political promise of higher future economic growth.

Secondly, we take the U.S. as a case study. The U.S. had major marginal tax rate reductions during the presidency of Ronald Reagan in the 1980s, and there was some tax simplification as well. After the 1980-1982 recession ended, economic growth was quite robust in the 1980s, although not quite as high as in the previous decade. The U.S. national debt went from 31% of nominal GDP in 1980 to 62% for 1992, as the tax cut experiment worked to worsen the finances of the U.S. federal government. Adding to the evidence was the impact of the modest tax increases in the 1990s, which did not appear to meaningfully hinder economic activity yet did dramatically improve government finances.

Finally, we take a look at the possible economic outcomes if the U.S. goes forward with large corporate and personal income tax cuts effective 2018. To preview our conclusions, we see little prospect for higher real GDP growth resulting from any tax reductions because we are pessimistic about tax simplification. We do see the potential for tax cuts assisting to sustain share values. We also see any tax cuts adding materially to the U.S. debt load. U.S. debt loads are moving into the territory that make the economy considerably more fragile, especially related to upward interest rate shocks. Increased fragility does not necessarily mean recession; however, fragility does increase the probabilities of a recession given a significant economic shock. In this scenario, tax cuts might lead to a much more cautious interest rate policy from the Federal Reserve, negatively impacting the U.S. dollar.

Figure 1: Laffer Curve (stylized) – top marginal tax rate versus tax revenue as percent of GDP



Source: Created as an illustration by CME Group Chief Economist

2. TAXATION THEORY AND THE "LAFFER CURVE"

The debate over the economic impact of tax cuts was energized back in 1970s with the work of Arthur Laffer, and became known in the political discourse in the Reagan years as supply-side economics. Arthur Laffer [Canto et al. (1982), Laffer (2004)], and various co-authors [Canto et al. (1982), Canto and Miles (1981)], produced some excellent research in this area. The theoretical model they developed linking tax cuts to future economic growth was both elegant and intuitive. The model also depended on some heroic simplifying assumptions; and as we can observe with hindsight, the devil was in the details of these unrealistic assumptions.

The essence of the relationship between tax rates and economic growth is intuitively visualized in the Laffer Curve. Starting from a zero top marginal tax rate, as tax rates rise, so do tax revenues as a percent of GDP – up to a point. That is, as the top marginal tax rate gets higher and higher, it ultimately serves as a disincentive

for individuals and corporations to seek higher earnings, and tax revenues as a percent of GDP start to fall even as the top tax rate goes higher and higher. Please note that the Laffer Curve is a stylized representation of the theory and the actual peak point of tax revenues related to tax rates is highly controversial, not to mention the shape of the curve itself. The Achilles heel of the Laffer Curve is the heroic assumption that the top marginal tax rate drives personal spending and corporate investment decisions. Unfortunately, tax codes are exceptionally complex and full of special deductions and loopholes. As a result, the link between the top marginal tax rate and actual consumer spending and business investment is tenuous, if non-existent.

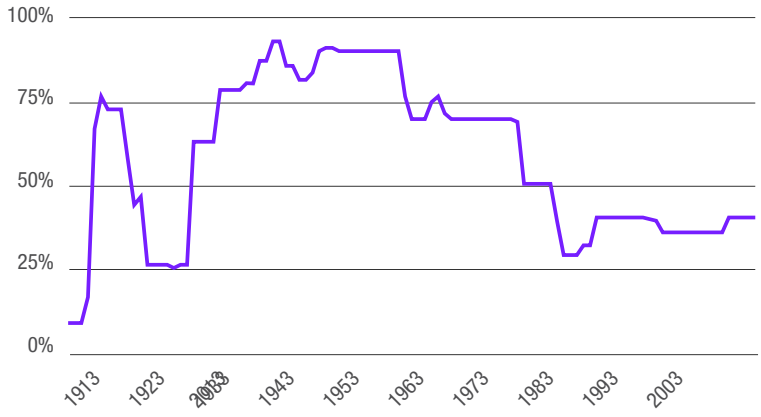
Moreover, even if corporations or individuals were to receive a large realized tax cut, there is little to guarantee that the cuts will impact the components of GDP. For example, corporations might decide to use the new-found money to buy back their stock, refinance their debt, raise dividends paid to shareholders, or make a strategic acquisition. While all of these activities have the potential to increase shareholder value, they do not contribute at all to real GDP growth. Only if corporations increase domestic business investment is there likely to be any link to future GDP growth. It works the same way for individuals, especially the wealthy. Wealthy individuals are much more likely to save more of their tax reduction than average wage earners. Hence, it matters in a significant way if the tax cut is tilted toward the wealthy or not. And then there is the question of whether government spending is held constant or not. If government spending is reduced to offset the short-term negative impact on budget deficits, then the actual spending from the tax cut will almost certainly not compensate for the reduced government spending due to part of the tax cut being saved by individuals or going for stock buybacks and dividend increases by corporations.

The strongest case for tax reform promoting economic growth is when there is meaningful tax simplification. Tax simplification opens the possibility for marginal tax rates to have more influence over economic decisions, since it would eliminate loopholes. Politically, there is often a lot of rhetoric about tax rate decreases being accompanied by tax simplification, but in practice it is exceedingly rare.

3. THE CASE OF THE U.S.

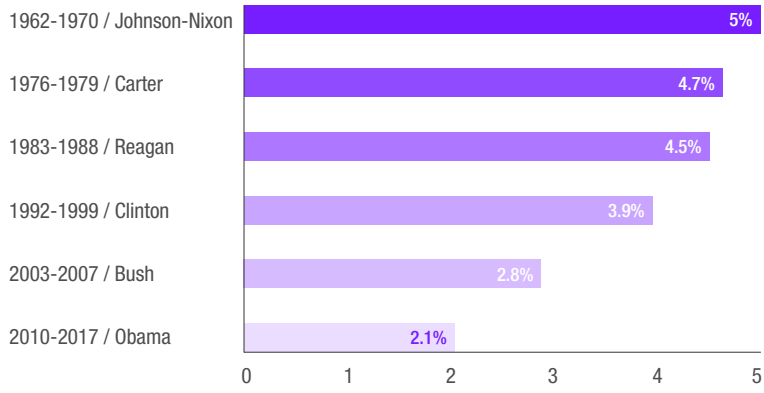
The U.S. has been a very interesting laboratory for analyzing tax changes. The personal and corporate income tax rate, as well as special deductions and loopholes, have been adjusted many times over the past century. Take the top marginal tax rate on personal income as an example. The rate started out in 1913 below 10% and applied to only the wealthiest of individuals. By the 1950s, the top rate was around 90%, but still applied to a relatively few. During the 1960s, the top tax rate was lowered to 70%, the tax base was broadened, and many deductions and loopholes were included in the tax code. The 1980s, under President Reagan, saw large cuts in the top rate, down to 28%,

Figure 2: Top U.S. marginal tax rate for personal income



Source: Tax Policy Center (http://tpc.io/2g2lETc)
Note: Top tax rate was applied to different income levels. Higher rates applied to only very few wealthy individuals. Lower tax rates typically broadened the tax base and applied to more people.

Figure 3: U.S. economic expansions – average annual GDP growth (recessions omitted)



Source: St. Louis Federal Reserve Bank FRED Database (GDPC1)

and some meaningful tax simplification, and the tax base for the top rate was considerably broadened. The 1990s, under Presidents Bush and Clinton, saw some increases in the top tax rate. In short, U.S. top tax rates have been all over the map and have applied to very different tax bases as well over time.

All of these tax rate changes did not have a discernible impact on the pattern of the last six decades of slowly decelerating economic growth rates. In the 1960s, the U.S. was arguably a 5% real GDP annual growth economy. Each decade since then, however, has shown a steady deceleration down to the 2% annual growth trend seen since the economic expansion following the Great Recession of 2008-2009.

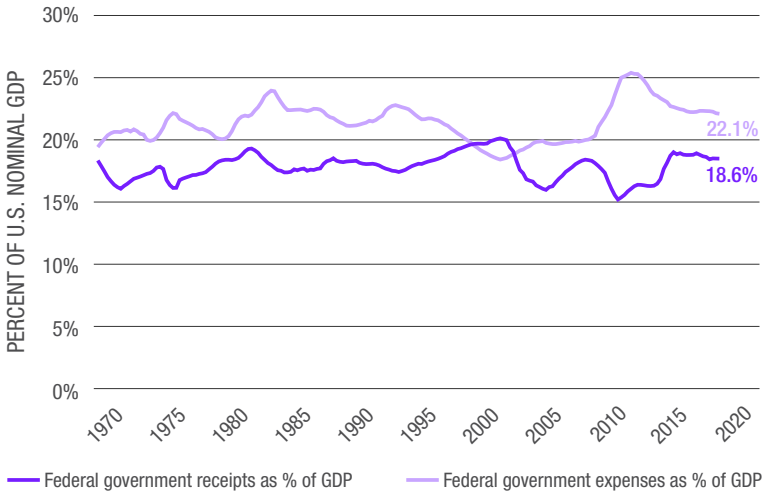
Indeed, we would argue that since 1950 the U.S. has experienced three growth drivers. The 1950s and 1960s were about recovery from the war, building a modern economy, and improving the infrastrucutre, such as the interstate highway system. The result was rapid growth in labor productivity and well above average GDP growth. The 1970s and 1980s were about the arrival of the large baby boomer generation into the workforce. The baby boomers, born after WWII and into the early 1960s, resulted in very rapid expansion of the labor force as they matured into their twenties in the 1970s and 1980s, keeping post War economic growth elevated, as these new workers were absorbed into the economy.

From the 1990s onward the drivers changed direction. The arithmetic is informative. Real GDP growth can be decomposed into growth in the labor force and growth in labor productivity. While labor productivity growth has ebbed recently, the demographic trend has been even more powerful, with aging boomers retiring and smaller generations following, leading to very low growth rates in the labor force – now below 1% in the U.S. Hence, it is hard to find an impact from all the different tax regimes, when demographic patterns explain such a substantial part of the deceleration of potential GDP growth.

Nevertheless, one cannot study the impact of tax rate cuts on GDP growth without special attention to the 1980s. In two stages, the highest marginal tax rate went from 70% to 28%, and there was some tax simplification. Economic growth in the 1980s held up very well as baby boomers were in their prime working years and not starting to leave the labor force yet. While tax revenues as a percent of GDP remained in the 17.5% to 18.6% zone during 1983-1990, budget deficits increased and the national debt soared. Indeed, the total outstanding U.S. national debt was about 31% of GDP in 1980 and was over 53% in 1990, at the end of the decade. The idea that lower tax rates, even with some tax simplification, would result in substantial additional GDP growth so that tax revenues would rise and close the budget deficit did not happen.

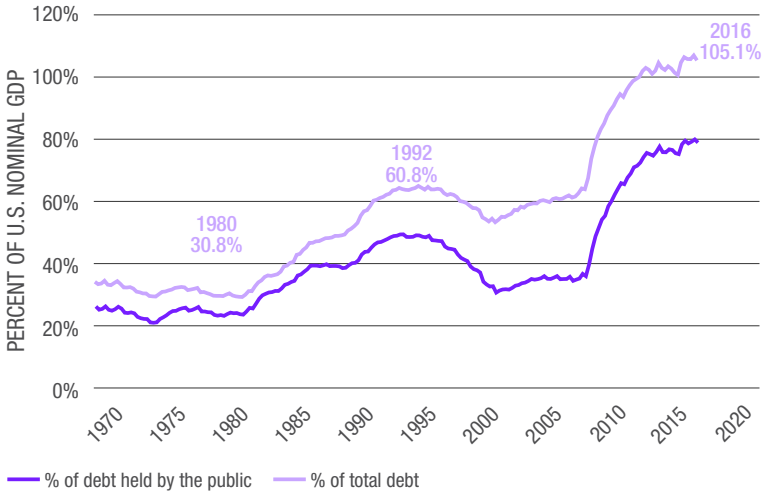
Subsequent U.S. Presidents in the 1990s, both Republican and Democrat, made the decision to close the budget gap with increases in tax rates. Federal debt fell from 64% of GDP in 1993 to 55% of GDP in 2001. This process was reversed in the 2000s, as government

Figure 4: U.S. federal government receipts and expenditures as percent of GDP



Source: St. Louis Federal Reserve Bank FRED Database (FGEXPND, FGRECPT, GDP)

Figure 5: U.S. national debt as percent of GDP



Source: St. Louis Federal Reserve Bank FRED Database (GFDEGDQ188S, FYGFGDQ188S)

expenses soared in the immediate aftermath of the Great Recession. By 2013, the national debt as a percent of GDP was 101% in the U.S. From 2013 through 2016, with tight controls over government spending and modest economic growth, national debt ratios remained relatively stable at just less than 105% of GDP.

4. LOOKING TO FUTURE U.S. TAX POLICY AND POTENTIAL MARKET SCENARIOS

U.S. tax policy featured prominently in the 2016 elections, and there is a strong likelihood of tax legislation making it through Congress and becoming effective in 2018. Even if the U.S. reduces both personal and corporate income tax rates, the devil will be in the details and probably not at all clear until the legislation passes both Houses of Congress – no mean feat, as we have observed during 2017 regarding the attempts to pass a new healthcare law.

The tax debate is likely to focus on three main challenges: (1) whether lower taxes will increase future economic growth as analyzed here, (2) the possibility of tax simplification, and (3) the implications for the budget deficit and national debt.

The debate about tax cuts and future economic growth will largely pit economists against politicians. Politicians in favor of a tax cut are going to consistently argue that higher growth will follow, allowing for lower projections of future deficits and national debt levels. Many economists, not all of course, will follow along the lines argued in this research, and will be very cautious in projecting higher economic growth in the face of severe demographic headwinds.

Our base case scenario assumes very little tax simplification because the special interest groups associated with each deduction and loophole are exceedingly strong. If one is willing to make the assumption of major tax simplification, then the case for stronger economic growth is much easier to make.

Estimating future budget deficits and national debt levels for the U.S. depend critically on the assumptions about growth that one is willing to make. And, politically, it will pit factions of the Republican party that have historically railed against rising the national debt versus other factions that want lower taxes no matter what and are willing to risk higher debt levels. It is this debt-versus-tax cut debate within the Republican Party that will largely determine the deals that need to be struck to pass tax cuts. Our base case scenario is that cuts in the personal income tax will be quite modest with little reform of current deductions, while we see much more consensus around a larger cut in corporate taxes. Hence, we will focus on some additional comments on corporate taxation in the U.S.

The first point to note is that corporate tax data in the U.S. is quite tricky to interpret. Let us take the 2016-2017 tax data as an example. Corporate tax receipts received by the U.S. federal government in the four quarters from July 2016 through June 2017 were reported as U.S.$409 billion. This headline number, used by many to analyze the size of the potential tax cuts, includes taxes paid by the Federal Reserve Bank, and that is a big problem for tax analysis.

Since the Great Recession, Federal Reserve Banks have gone through a period of massive asset purchases (i.e., quantitative easing or QE), and they earn substantial profits on their huge portfolios of U.S. Treasury securities and mortgage-backed securities. After keeping a very small surplus to add to their paid-in capital, the Federal Reserve Banks make quite large contributions to the U.S. Treasury. From July 2016 through June 2017, these contributions, reported as corporate taxes paid by the Fed, totaled U.S.$86 billion, on earnings of U.S.$87, or an effective rate of almost 99%. Clearly, the Federal Reserve is a special case and should be excluded from an analysis of U.S. corporate taxes.

Figure 6: U.S. federal corporate taxes and profits as percent of GDP

Source: St. Louis Federal Reserve Bank FRED Database. Corporate Profits = A053RC1Q027SBEA, Federal Reserve Profits = B397RC1Q027SBEA, Federal Corporate Taxes = B075RC1Q027SBEA, Federal Reserve Bank Taxes = B677RC1Q027SBEA, US Nominal GDP = GDP. Note: Corporate profits and taxes are net of Federal Reserve Bank profits and taxes.

The current annual level of corporate taxation, not including the Fed, is about U.S.$323 billion, or about 1.7% of GDP. Put another way, the scope for corporate tax cuts is not all that large in terms of a percentage of GDP, and if most of the tax savings are expected to go toward stock buybacks, dividend increases, debt pay-downs, and acquisitions, then one can easily see the benefits to stock prices, just not for the economy.

Also worth noting is the role of tax loopholes. U.S. corporate profits on a GDP basis before taxes (and not including earnings of Federal Reserve Banks) were running at 11.35% of GDP for the Q3/16 – Q2/17 period.[2] As noted earlier, corporate tax receipts, not including the Federal Reserve tax payments, ran at 1.7% of GDP over the same period. Given that the top marginal rate is 35%, it is clear that corporations mostly do not pay the top rate, as that would imply corporate tax receipts of almost 4% of GDP. Indeed, the average effective corporate tax rate is about 15%. What this means is that is that top tax rate does not drive business investment decisions, and so expectations of increased capital spending in the U.S. from a cut in the top marginal tax rate are not easily justified.

Given our relatively dim view that tax cuts will increase GDP growth, our perspective is that the budget deficit will rise, increasing the national debt as a percent of GDP. This scenario is associated with two very distinct possible market impacts.

First, tax cuts are bullish for stocks. How much the chances of a tax cut is discounted in the current share prices is highly debatable. However, given the legislative hurdles, it would seem that some discount is being applied to tax cut probabilities, so any tax cut might have a further positive impact on shares.

What is less obvious is the possible impact on Federal Reserve interest rate policy. With inflation stuck just below 2% and our base-case scenario not assuming much above 2% real GDP growth, the Federal Reserve may adopt a bias in 2018 after tax legislation is passed, if it is passed, that the possibility of increases in the national debt make the economy more fragile and argue against higher interest rates (i.e., higher interest expense). That is, this base-case scenario sees tax cuts and a higher national debt associated with a more dovish Federal Reserve, which has the potential to put downward pressure on the U.S. dollar.

Of course, if one rejects our base case and goes with a higher-growth scenario, the market expectations are quite different. Higher-growth expectations argue for higher short-term interest rates and a very different path for the Federal Reserve, and probably for the U.S. dollar. Further, while equities are still beneficiaries of any tax cut, not all shares may respond in the same manner. If the tax legislation includes closing some loopholes in the corporate tax code, this is expected to hit mega-tech and large-company stocks harder than small-company stocks. Hence, if one adopts the loophole-closing scenario for the tax cut, the Russell 2000 small company index might do better than the mega-tech heavy NASDAQ 100 index.

[2] Source: St. Louis Federal Reserve Bank FRED Database. Corporate Profits before tax = A053RC1Q027SBEA, Federal Reserve Profits = B397RC1Q027SBEA, Federal Corporate Taxes = B075RC1Q027SBEA, Federal Reserve Bank Taxes = B677RC1Q027SBEA, US Nominal GDP = GDP).

References

Canto, V. A., D. H. Joines, and A. B. Laffer, 1982, Foundations of supply-side economics – theory and evidence, Academic Press

Canto, V. A., and M. A. Miles, 1981, "The missing equation: the wedge model alternative," Journal of Macroeconomics 3:2, 247-269

Laffer, A. B., 2004, "The Laffer curve: past, present, and future," Heritage Foundation Backgrounder 1765, 1176-96

Actively managed versus passive mutual funds: A race of two portfolios

ATANU SAHA | Chairman, Data Science Partners[1]
ALEX RINAUDO | Chief Executive, Data Science Partners

ABSTRACT

This paper demonstrates that the average investor would be better off by following a readily-implementable strategy of investing in a portfolio of the five largest active funds in U.S. equity, fixed income, and international equity asset categories than investing in a corresponding portfolio of passive index funds. The active-fund-portfolio outperforms not only in terms of average returns, but also in risk-adjusted returns, providing far greater downside risk protection than the passive fund portfolio. This paper has important implications for investment advisors because its findings question the "wisdom" of index investing, which has been receiving considerable attention in the financial press in recent years.

[1] Akhil Shah and Sonya Rauschenbach have provided invaluable research assistance for this article. We are immensely grateful to Burton Malkiel for his comments on an earlier version of the paper. Data Science Partners received no sponsorship or funding for this research.

1. INTRODUCTION

Many academic studies have claimed, and the financial press have touted, the benefits of passive index funds over actively managed mutual funds. The thrust of the argument is that passive funds provide superior performance relative to their actively managed counterparts primarily because of their lower fees. The objective of this study is to examine this argument from the perspective of an investor.

On April 6, 2016, the U.S. Department of Labor unveiled the final version of a rule designed to ensure that advisors who help customers invest in 401(k), Individual Retirement Accounts, and other retirement plans are putting their customers' interests ahead of their own when it comes to investment products, including mutual funds. In light of this rule, which takes effect in January 2018, the issue of the selection of an appropriate mutual fund has assumed greater significance.

The advice that investors are better off by sticking to a buy-and-hold strategy of investing in passive index funds has dominated academic studies for some time. For example, in a widely-cited study, Malkiel (1995) counseled: "Most investors would be considerably better off by purchasing a low expense index fund than by trying to select an active fund manager who appears to possess a 'hot hand.' Since active management generally fails to provide excess returns and tends to generate greater tax burdens for investors, the advantage of passive management holds, a fortiori."

The "wisdom" of index investing has also been widely covered in the media, particularly in recent times. In August 2014, a Market Watch article asked: "The theoretical benefits of active management have proven to be fables. So why are investors still paying high fees for disappointing, inconsistent and tax inefficient performance?" [Sisti (2014)]. In 2015, the Wall Street Journal commented: "Index funds don't just outperform most actively managed mutual funds. They also make more money for investors" [Clements (2015)].

Morningstar, a leading mutual fund research firm, in a June 2015 report stated: "Actively managed funds have generally underperformed their passive counterparts, especially over longer time horizons and experienced high mortality rates (i.e. many are merged or closed). In addition, the report finds that failure tended to be positively correlated with fees (i.e. higher cost funds were more likely to underperform or be shuttered or merged away and lower-cost funds were likelier to survive and enjoyed greater odds of success)" [Johnson et al. (2015)].

Investors appear to be listening to the media coverage of the benefits of index investing.[2] For example, 2015 saw record inflows into passive mutual funds and record outflows from actively managed funds. On January 14th, 2016, the Wall Street Journal reported that in 2015 investors removed U.S.$207.3 billion from U.S. actively managed mutual funds, while index funds received an inflow of U.S.$413.8 billion. The same article concluded: "The outflow represents a stark change in investor attitude toward equities as investors wrestle with new stresses in a bull market that has lasted nearly seven years" [Krouse and Driebusch (2016)]. A relatively recent CNBC article commented: "Pity the active fund manager. More dollars have flowed to index strategies that track a market benchmark, such as the S&P 500 index, partly because such funds typically have lower costs than active funds and more investors believe that stock-picking managers can't regularly beat the financial markets" [Anderson (2015)].

"The outflow represents a stark change in investor attitude toward equities as investors wrestle with new stresses in a bull market that has lasted nearly seven years."

[Krouse and Driebusch (2016)]

Yet, a sizeable portion of mutual fund assets continue to reside in actively managed funds. According to Morningstar's 2014 Annual Global Flows Report, passive index funds hold only 24% of the total asset under management (AUM) for U.S. mutual funds [Lamy and Strauts (2015)]. However, this share is markedly higher than what it was ten years ago; another Morningstar article reported that at the end of 2003 only 12% of assets of all U.S. open-ended mutual funds were in passive funds [Zoll (2014)]. This pronounced rise in the AUM-share of passive funds potentially reflects two factors: (a) fund flows, with investors increasingly choosing passives over actives in recent years; and (b) the possible relative outperformance of the passive funds, because the growth of any fund's AUM is affected not only by fund flows but more importantly by its performance. As this paper discusses in more detail, data do corroborate the first factor; examining the veracity of the second is the focus of this study.

[2] Sirri and Tufano (1998) suggested there is some relationship between media coverage and fund flows.

2. THE PRIOR LITERATURE

A large body of literature has examined the performance of mutual funds. These studies can be categorized into two groups. The ones in the first group have analyzed the performance of actively managed funds relative to the relevant market indices. A subset of these studies has also explicitly evaluated the adverse effects of fees on fund performance. The second strand of literature, which is considerably smaller than the first, has investigated the comparative performance of active and passive mutual funds.

2.1 Studies comparing active funds' performance to benchmark indices

Many prior studies have judged the performance of actively managed funds by comparing their returns to a benchmark index. One of the earliest in this body of literature is the study by Jensen (1968), which used data for the period 1945–1964. He found that a majority of mutual funds, which, at that time, were mainly actively managed equity funds, generated a negative alpha, that is, they underperformed the market after accounting for systematic risk of the fund's returns.

A number of studies have built on Jensen's research. Augmenting the three-factor model in Fama et al. (1993), Carhart (1997) analyzed mutual fund performance using a four-factor model. His study, using equity mutual funds' data from 1962–1993, showed that funds' net returns are negatively correlated with expenses and that the more actively the manager trades, the lower the net return, i.e., return net of fees. Fama and French (2010) used both their three-factor and Carhart's four-factor models to examine active mutual funds' performance. They used data from the period 1984–2006 to create a portfolio of NYSE, Amex, and NASDAQ stocks; they then compared the performance of actively managed U.S. equity funds to this benchmark portfolio. They concluded that the majority of actively managed funds do not generate superior returns relative to the benchmark primarily due to their high fees.

Other studies have assessed both gross and net returns of actively managed mutual funds to specifically determine the impact of fees on fund performance. Malkiel (1995) examined equity mutual funds over the period 1971–1991, comparing their performance to the Wilshire 5000 and the S&P 500 market indices. He found that actively managed funds underperformed these benchmarks both before and after expenses, thus concluding that high fees alone cannot explain their underperformance. Wermers (2000) analyzed the stock selection acumen of fund managers using data for the period 1975–1994 and found that, on average, these managers' selections outperformed the market benchmark by 1.3% annually. However, their funds' returns net of fees underperformed the market by 1%. Grinblatt and Titman (1989, 1993) showed that certain active fund managers outperformed benchmarks before expenses, especially managers of aggressive growth funds. However, these funds also had the highest expenses; as a result, their performance, net of expenses, lagged their benchmarks.

Notwithstanding the general finding that actively managed funds do not outperform benchmarks, prior research has also shown that in certain market segments and business cycles, actively managed funds can benefit investors. For example, Fortin and Michelson (1999) found that over the period 1976–1995 small-cap active funds outperformed the Russel 2000 index. The authors suggested that inefficiencies in the small-cap market segment allow for more potentially skilled stock pickings. Kacperczyk et al. (2005) showed that between 1984 and 1999 active equity funds that made concentrated industry bets usually beat their benchmark portfolios. They concluded that fund managers' stock-picking abilities are likely to be more evident in industries they specialize in.

2.2 Studies comparing active and passive funds' performance

While many studies have compared actively managed funds to benchmark indices, relatively few have compared the performance of actively managed funds to their index fund counterparts. A possible reason for the paucity of such studies is that much of the prior research on mutual fund performance has relied on data from years when index funds were either nonexistent or a relatively new phenomenon. The relatively few recent studies that have compared actively managed funds to their indexed counterparts have found mixed results on comparative performance.

Fortin and Michelson (2002) compared the performance of actively managed funds to Vanguard index funds that they deemed to be the appropriate counterparts. This study not only confirmed the finding in their 1999 paper (relative outperformance of actively managed small-cap funds to the Russel 2000 index) but also found that international stock funds outperformed the corresponding Vanguard index funds. Interestingly, their 2002 paper also found that actively managed funds

tend to outperform index funds when the economy is transitioning into or out of a recession; in particular, this outperformance was most pronounced during the years: 1979-82, 1991-93, and 1999-2000. Based on this finding they concluded: "It appears that active fund management is better than index funds at guiding portfolios through rough times."

Holmes (2007) compared the performance of actively managed and index funds using data for the period 1995–2004, and segregated the comparative analysis based on the Morningstar asset categorization of funds. Her results were mixed. Actively managed funds in the asset categories of mid-cap value, small-cap blend, and international mid/small cap blend outperformed their respective index funds; however, index funds outperformed in all large cap asset classes, U.S. mid-cap blend, and small-cap value and growth asset categories. She also compared actively managed funds' performance to their respective S&P market segment indices and found that actively managed funds in all asset classes underperformed the market indices. However, she observed that active funds fared better during the 2000-2002 market downturn, which is consistent with Fortin and Michelson's (2002) results.

This study builds upon the foregoing strand of literature that compares active and passive funds' performance. Our paper does not compare the performance of active funds to market indices because these indices are not investment options for investors, but passive funds, which track the indices, are. The realized returns to an investor in an index fund can be different, sometimes not insubstantially, from the returns of the underlying index because of tracking error and fund fees. This paper, therefore, compares the investment performance, net of fund fees, of actively managed funds with that of asset category-matched passive funds. The principal contribution of this paper is the comparison, from an investor's perspective, of the relative performance of two portfolios of active and passive funds based on an implementable investment strategy. This paper begins by investigating the relative sizes and fund flows of active and passive funds.

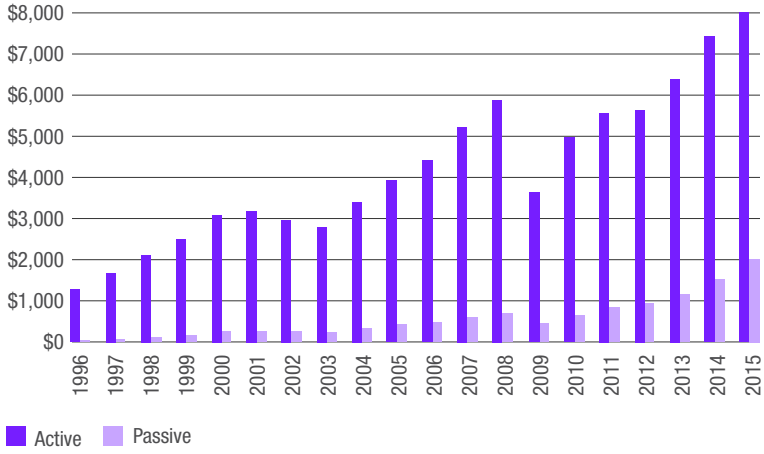
3. AUM AND FUND FLOW ANALYSIS

This study is based on data from Morningstar's open-end U.S. mutual fund database. Data was downloaded for funds that are categorized as having investment focus in three broad asset categories: U.S. equity (including sector funds), international equity, and fixed income (including both taxable and municipal funds). The dataset is comprised of 77,687 fund-year observations across 7,469 unique funds; of these, 7,155 were actively managed and 314 were passive index funds. The dataset is free from survivorship bias because it encompasses all funds, dead or alive, during the entire period, 1996-2015. Further details regarding the cleaning and organization of the dataset are available from the authors.

This section first discusses the size of aggregate assets in active and passive funds and then examines the data on fund flows. The findings corroborate the commentary in the financial press regarding the investors' increasing choice of index funds, particularly in recent years.

Figure 1 shows that the AUM for active funds has grown from U.S.$1.3 trillion in 1996 to nearly U.S.$8 trillion in 2015. By contrast, passive funds' AUM exceeded U.S.$1.3 trillion for the first time in 2014. Notwithstanding the growth of passive index funds, and perhaps due to inertia or lack of passive funds in certain retirement accounts, a vast majority of investors' assets continue to reside in actively managed funds.

Figure 1: AUM of active and passive mutual funds (U.S.$ bln)



Source: Morningstar

However, the trend in the relative share of the two types of funds is unmistakable: 20 years ago, passive funds had only 3% of the AUM all funds and by the end of 2015 this figure has grown to 20%. The growth in the share of assets garnered by passive funds reflects two related yet distinct factors, fund flows and performance.

The first factor is examined in Figure 2, which shows the net flows as a percent of assets each year for

both active and passive funds from 2000 to 2015.[3] In the past nine years, with the exception of 2009, the inflow percentages in passive funds have remained higher than those of active funds. In fact, in these nine years, active funds have had only three years of positive inflows. These findings are consistent with the commentary in the financial press discussed in the introduction of the paper.

4. FUND FEES AND RELATIVE PERFORMANCE

The popular press has suggested that one of the primary reasons for the observed pattern of fund flows is higher fees charged by the actively managed funds. The academic literature also attributes high fees as the principal reason for the underperformance of these funds [Jensen (1968), Malkiel (1995), Carhart (1997), Fama and French (2010), Grinblatt and Titman (1989), Grinblatt and Titman (1993), and Wermers (2000)]. This section initially examines the relationship between fund fees and the performance of active funds. This is followed by an examination of the relative performance, net of fees, of all active and passive funds in the dataset.

Morningstar provides data on each fund's expense ratio, which is the percent of a fund's assets used to pay for its operating expenses and management fees, including 12b-1 fees.[4] The weighted average expense ratios have been computed for all active and passive funds in the dataset for each of the twenty years. The weights are each fund's annual AUM expressed as a ratio of all funds' total AUM in their respective management style category of active or passive. Specifically, denoting *w* as the weight, *A* as the AUM, and *N* as the number of funds:

$$w_{it}^{j} = \frac{A_{it}^{j}}{\sum_{i=1}^{N} A_{it}^{j}}, \quad i = i^{th} \text{ fund; } j = \text{active, passive; } t = t^{th} \text{ year} \qquad (1)$$

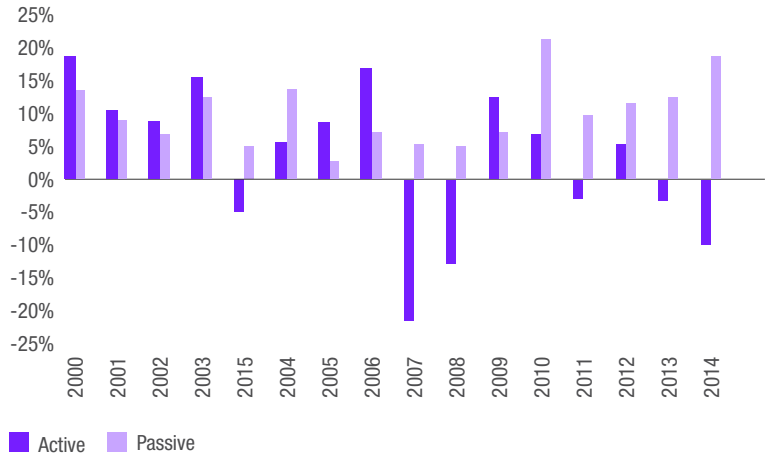
Denoting the expense ratio as *er*, the weighted average expense ratios were calculated as:

$$\overline{er_t} = \sum_{i=1}^{N} w_{it}^{j} \times er_{it}^{j} \qquad (2)$$

Figure 3 shows the asset-weighted average expense ratios for active and passive funds. As is evident from Figure 3, passive funds have considerably lower fees than active funds, although both categories of funds show a slight downward trend since 2002.

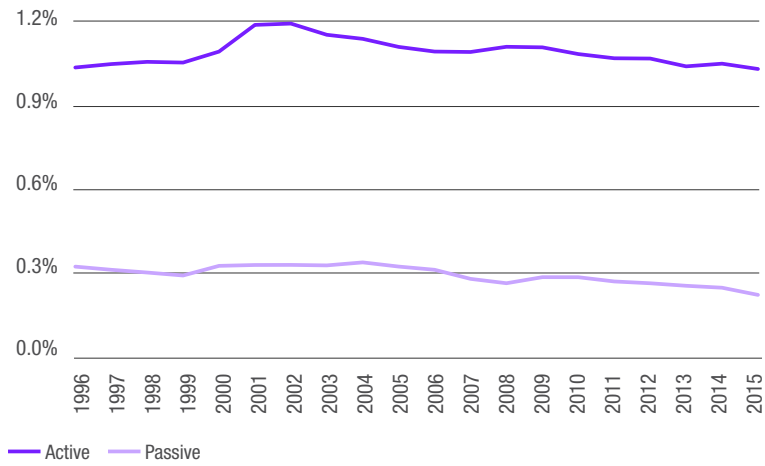
The fact that active funds have higher fees does not necessarily mean that they are bad investments.

Figure 2: Annual fund flows as % of AUM



Source: Morningstar

Figure 3: Expense ratios over time



Source: Morningstar

Ultimately, investors care about returns net of fees. Since fees are deducted from a fund's assets, one might be tempted to argue higher fees lead to lower returns for active funds than passives. However, that is not necessarily the case. It is possible for a fund to charge high fees but also deliver above average returns such that its net-of-fee returns are higher than returns offered by a lower-fee fund.

[3] Morningstar data on fund flows are sporadic, and possibly unreliable, prior to 2000.
[4] Morningstar defines expense ratio as follows: "The percentage of fund assets used to pay for operating expenses and management fees, including 12b-1 fees, administrative fees, and all other asset-based costs incurred by the fund, except brokerage costs. Fund expenses are reflected in the fund's NAV" (net asset value).
[5] All fund returns data used in this paper are returns net of fees and expenses. Also, all returns are logarithmic returns.

Table 1: Regression results

EXPLAINED VARIABLE: AVERAGE ANNUAL FUND RETURNS (T-STATS IN PARENTHESIS)					
	A	B	C	D	E
Expense ratio (ER)	-2.1954 (-12.88)			-4.3049 (-12.97)	0.5306 3.63
No. of years		0.0034 (22.87)		-0.0003 (-0.82)	
Fund size (in U.S.$ mln)			1.9371 (5.74)	0.4510 (1.36)	
ERxNo. of years				0.2529 (9.24)	
Intercept	0.0673 (25.85)	0.0011 (0.58)	0.0350 (33.35)	0.0646 (12.49)	0.0517 (27.68)
Adjusted-R^2	0.0225	0.0680	0.0045	0.0912	0.0098
Number of observations	7,155	7,155	7,155	7,155	1,239

The relationship between the net-of-fees returns[5] and fees for active funds are examined through cross-sectional regression analyses. The results of the regression analyses are shown in Table 1. In each of the models (models A-E), the explained variable in the regression is the average annual returns of each fund, which is computed over the years of its existence in our dataset. In model A, the explanatory variable is the fund's expense ratio. In model B, the explanatory variable is number of years of fund data, which reflects the number of years the fund has been in existence, i.e., fund-life. In model C, the explanatory variable is fund size, which is captured by the average AUM of the fund. In model D, in addition to the preceding explanatory variables, an interaction variable between expense ratio and fund-life is introduced.

In the first regression (model A), the estimated coefficient for the expense ratio variable is negative and statistically significant, suggesting that higher expenses are, on average, associated with lower returns for active funds; consistent with the findings in the prior literature.

The results of models B and C show that both fund-life and fund size are associated with higher returns for active funds. In model D, the coefficient of the interaction variable, between expense ratio and fund-life, is positive and statistically significant. This result suggests that the negative effect of fees is outweighed by superior performance for funds with longer life. This finding jibes with intuition because funds that have been in existence for many years are likely to be ones that have delivered good performance over the years.

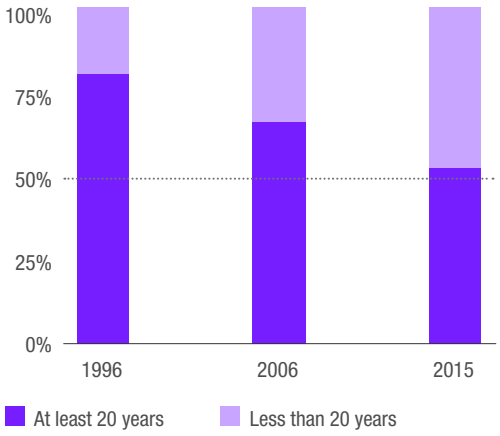
In model E, this issue is further examined by limiting the sample for the regression analysis to the 1,239 funds that have been in existence over the entire twenty years of our dataset. In this regression, the relationship between performance and expense ratio switches sign, becoming positive and statistically significant. This result is consistent with the finding from model D, where the coefficient of the interaction variable was found to be positive. Of course, this subset of 1,239 actively managed funds are the survivors: it is unlikely that these funds would have survived 20 or more years without delivering superior performance, net of fees.

While these survivors constitute a fairly modest percent of the number of actively managed funds, they control a significant portion of assets. As shown in Figure 4, in 1996 they had nearly 80% of assets of all actively managed funds; in 2015 more than half of all active fund assets resided in these funds.

4.1 Relative performance

This paper now turns its attention to the comparison of the performance of actively managed and passive index funds. For this analysis, data on all funds is

Figure 4: Percent of AUM for active funds existing at least 20 years



Source: Morningstar

used regardless of the number of years of existence. The question this study seeks to answer is whether, on average, actively managed funds provide superior performance relative to passive index funds. Conditioning this comparative analysis on a certain minimum number of years of a fund's existence would have introduced survivorship bias.

The funds are segregated by asset category, based on Morningstar's categorization. For each of the three asset categories (U.S. equity, fixed income, and international equity) and for each of the two management styles (active and passive) the weighted average annual returns are computed, where the weights are each fund's annual AUM as a ratio of all funds' AUM in that asset category and management style for that year.

Specifically, for each of the three asset categories, the weighted average annual returns, denoted by r, are calculated as:

$$\overline{r_t}^{j} = \sum_{i=1}^{N} w_{it}^{j} \times r_{it}^{j}\ ,\ i = i^{th} \text{ fund; } j = \text{active, passive; } t = t^{th} \text{ year} \quad (3)$$

where the weights in equation (3) are given by the expression in equation (1), with the exception that these weights are computed separately for each of the three asset categories.

The performance metrics shown in Table 2 are computed from the weighted average annual returns over the twenty-year period, 1996-2015.

Table 2 shows that active funds, on average, underperform their passive fund counterparts in the U.S. equity and fixed income categories. This underperformance is also evident on a risk-adjusted basis, as seen from the Sharpe and Sortino ratios. For fixed income funds, the Sortino ratio, which measures downside risk, is markedly superior for passive funds. These results, however, do not carry over to international equity funds.[6] Both in absolute and in risk-adjusted returns, actively managed funds, on average, outperform passive funds. One possible explanation for this result could be that fund managers' acumen and research play an important role for investment choices in foreign equities. This result is consistent with similar findings in the prior literature that found outperformance of active funds in certain market segments, such as mid-cap value, small-cap blend, and international mid/small cap blend [Fortin and Michelson (2002)].

Table 2: Performance metrics for all active and passive mutual funds: 1996-2015

	U.S. EQUITY		FIXED INCOME		INTERNATIONAL EQUITY	
	ACTIVE	PASSIVE	ACTIVE	PASSIVE	ACTIVE	PASSIVE
Mean returns	6.4%	7.7%	4.2%	4.9%	5.3%	3.7%
5th percentile	-27.2%	-25.6%	-1.6%	-1.0%	-20.9%	-24.9%
Standard deviation	19.6%	19.3%	5.0%	3.3%	22.0%	22.3%
Sharpe ratio	0.21	0.28	0.37	0.75	0.13	0.06
Sortino ratio	0.13	0.18	0.25	1.37	0.08	0.04
Number of funds	3,560	220	2,372	38	1,223	56

[6] Note that the difference in mean returns were not statistically significant for any category.

5. A HORSE RACE OF TWO PORTFOLIOS

Based on the results in Table 2, one might be tempted to conclude that investors should avoid actively managed funds, particularly in the U.S. equity and fixed income asset categories. However, it would be injudicious to jump to this conclusion without considering two important issues. First, Table 2 reflects the average performance of thousands of funds. As a result, these findings provide little guidance as to how one would go about choosing a fund or a set of funds to invest in. Expressed differently, because average performance metrics do not provide an implementable investment strategy, one cannot objectively determine how a typical investor's portfolio of active or passive funds would have performed over time. Second, most investors hold a diversified portfolio of funds, allocating investments across asset categories, such as U.S. and international equity, fixed income, etc.[7] Consequently, to evaluate the relative performance of active and passive funds one must account for the relative weights of these asset categories (U.S. equity, international equity, fixed income) in investors' portfolios.

The horse race of active and passive fund portfolios based on actual historical returns addresses both these issues. Specifically, two investable portfolios of active and passive funds are constructed adopting the following steps:

Step 1: for each management style (active, passive), equally-weighted portfolios of the five largest funds (by AUM as of November 30th of the prior year) are created in each of the three categories: U.S. equity, fixed income, and international equity. So, each year, the two portfolios of active and passive funds have 15 funds each, corresponding to the five largest funds in the three asset categories.

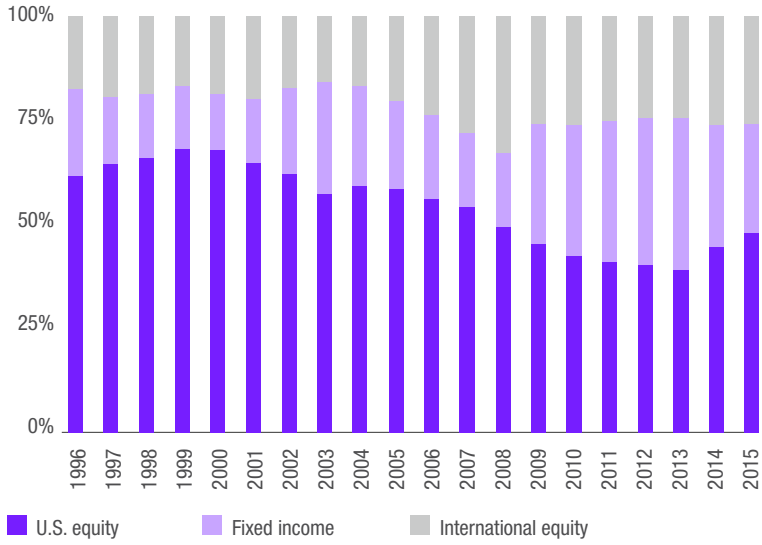
Step 2: each year, the returns of the fifteen funds are then combined to a single portfolio return of either active or passive funds, using asset category weights. These asset category weights are computed as follows: first, in any given year *t*, the aggregate AUM of funds, regardless of management style or asset category, is computed by summing the AUMs of all 30 funds; let this aggregate AUM be denoted by AS_t. Then the asset category weight, denoted by *aw*, is calculated as follows:

$$aw_{kt} = \frac{\sum_{i=1}^{10} S_{it}^{k}}{AS_t}, \quad i = i^{th} \text{ fund; } t = t^{th} \text{ year; } k = k^{th} \text{ asset category} \quad (4)$$

where *S* denotes the AUM of each fund in each asset category. Note in (4), the summation in the numerator is across all top ten funds (five actives and five passives) in each asset category; and the denominator is the combined AUMs of all 30 funds. Thus, in any given year, the asset category weights are the same for active and passive funds.

This weighting scheme was chosen because in the early years of the sample period a vast majority of the passive funds were U.S. equity funds; as a result, had weights based on management style been used, the portfolio of passive funds would have received disproportionately higher weights for U.S. equity in the early years, especially through 2007. This weighting scheme would have merely reflected the fact that far fewer passive non-U.S. equity funds existed in the 1996-2007 period, and would not have captured investors' actual asset allocation choices.

Figure 5: Weighted portfolio



Source: Morningstar

It important to note that these asset category weights were determined by funds' AUM data and are not arbitrarily assigned (such as 60% equity, 40% fixed income, etc.) in the construction of the portfolios for the horse race. These weights, therefore, reflect both the investor choices and performance of funds for the three asset categories. Additionally, so as to not create a bias in the comparison of active and passive funds,

[7] These are top-level asset categories from Morningstar. Other equity funds invest in both developed and emerging market equities.

the same asset category weights are applied to both portfolios. The yearly asset category weights computed using (4) above are shown in Figure 5. It shows, notwithstanding the slight uptick in the last two years, that there has been a gradual shift away from U.S. equity to the other two categories, especially during 2001-2013. Over the entire 20-year period, the average weights were: U.S. equity 55%, fixed income 23% and international equity 22%.

Once the asset category weights have been determined, the 15 funds' annual returns are aggregated to a single portfolio return of either active or passive funds using those weights as follows:

$$Pr_{jt} = \sum_{k=1}^{3} a\, w_{kt} \times \overline{r_{jt}}^{k}\,,\ j = \text{active, passive,}\ k = \text{asset category} \quad (5a)$$

where,

$$\overline{r_{jt}}^{k} = 1/5 \sum_{i=1}^{5} r_{ijt}^{k}\,,\ k = 1, 2, 3 \quad (5b)$$

Thus, in (5a), *Pr* denotes the asset-category weighted annual returns of the portfolio of either active or passive funds.

Table 3 contains the results of this real-world horse-race; it shows the annual returns of the two portfolios

Table 3: Performance metrics for asset category – weighted portfolios

	ACTIVE	PASSIVE	LOST U.S.$ MLN
1996	12.8%	14.6%	-544
1997	18.8%	19.3%	-234
1998	17.4%	19.5%	-1,783
1999	14.0%	17.3%	-4,150
2000	-2.8%	-9.3%	12,312
2001	-8.9%	-11.6%	5,013
2002	-14.9%	-16.6%	3,215
2003	19.2%	21.0%	-3,116
2004	9.3%	10.3%	-2,315
2005	8.2%	6.4%	4,733
2006	11.9%	14.5%	-8,032
2007	9.3%	8.3%	3,973
2008	-41.1%	-42.0%	3,915
2009	22.9%	21.5%	4,397
2010	9.7%	11.9%	-9,300
2011	-2.9%	-1.7%	-7,149
2012	13.7%	11.6%	13,446
2013	15.2%	11.4%	29,426
2014	4.9%	6.0%	-10,888
2015	0.5%	-1.5%	27,298
			60,218
Mean returns	5.9%	5.5%	
5th percentile	-16.2%	-17.9%	
Standard deviation	14.8%	15.7%	
Sharpe ratio	0.24	0.20	
Sortino ratio	0.13	0.12	

of active and passive funds for the years 1996-2015. The bottom part of this table also contains the relevant performance metrics.

The results in Table 3 show that the portfolio of top 15 active funds outperformed the top 15 passive index funds, both in terms of average returns and risk adjusted returns.[8] Importantly, the active portfolio also provided superior downside risk protection as is seen by the better Sortino ratio; this is further corroborated by the active portfolio's outperformance in years the market experienced severe downturns, 2000-2001 and 2008.

Consistent with the performance numbers shown in Table 2, the performance difference between the active and passive fund portfolios is largely driven by the outperformance of active funds in the other-equity category. As noted earlier, the other equity category includes both developed and emerging market funds. We explored the impact of removing emerging markets funds from the category of other equity. Specifically, we reconstructed the other equity portfolios excluding those funds that were exclusively emerging market funds, and reran the horse race, keeping the funds in the U.S. equity and fixed-income categories unchanged. While this reconstruction does reduce the difference between the mean returns of the active and passive portfolios over the 1996-2015 period by 0.08% (i.e., the annual average return difference drops from 0.31% to 0.23%), our main conclusion still holds: the active portfolio outperforms the passive portfolio.

The economic implications of the horse-race results are also illustrated in the last column of Table 3; it shows the incremental sum investors would have gained had they invested in the top-15 active funds as compared to the top 15 passive funds. These figures are computed by multiplying the annual differences in the two portfolio's returns by the total assets in the top-15 passive funds each year. As shown in the bottom of the column, over the 20-year period this difference sums to over U.S.$60 billion; however, a sizeable portion of this difference, U.S.$59 billion, occurs in the last four years. While it is unlikely that the assets in the top 15 passive funds could actually be redeployed to the top 15 active funds without impacting fund performance, the result shown in Table 3 illustrates the economic impact of small performance differences.

5.1 A real-world illustration

Table 3 shows that the average annual returns of the active and passive portfolios are 5.9% and 5.5%, respectively. This difference of 0.4%, while small, can have a non-trivial effect on an investment account balance over time. This is illustrated through a simple exercise in this sub-section. It is assumed that two individual retirement accounts (IRA) start with the identical sum of $10,000 in 1995; one IRA account invests in the largest active funds discussed in the horse-race, while the other invests in the largest passive funds. Additionally, it is assumed that each account receives the maximum allowable IRA contribution at the beginning of each year. The annual returns each account would experience between 1996 and 2015 are listed in Table 3. The performance of the two IRA accounts over time is shown in Figure 6.

Figure 6: Investment balance in a portfolio of active versus passive funds

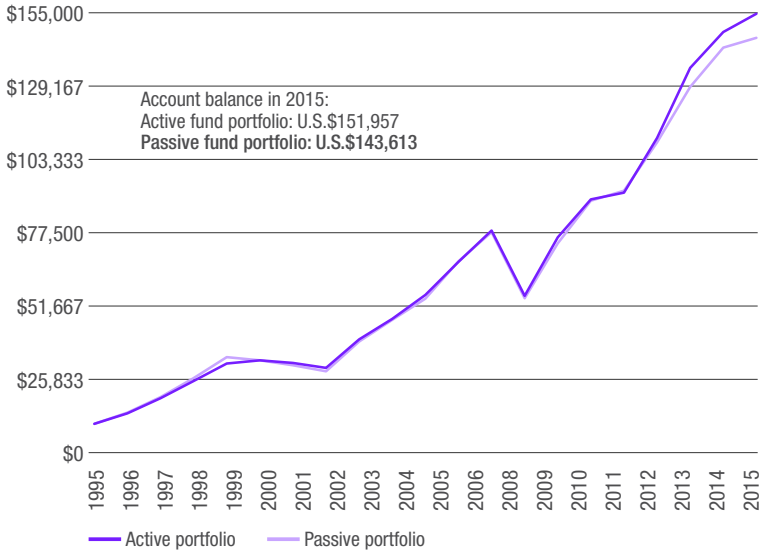


Figure 6 shows that the account comprised of the active funds cumulatively outperforms the passive one over the twenty-year period. By 2015, the active fund portfolio has an account balance of approximately $152k dollars, while the passive fund portfolio's balance is about $144k. Additionally, it is also found that in 14 of the 20 years, the active fund account balance is higher, despite the fact that in the horse-race the active portfolio's return is superior only in 10 years. It is important to note that, up through 2015, the cumulative performance of the active fund is higher than that of the passive one regardless of the start year of the horse-race.

[8] These results hold even when, in constructing the horse race portfolios, fixed weights of 55%, 23%, and 22% are used across all years, for the three asset categories, U.S. equity, fixed income, and international equity. These weights are average category weights over the period 1996-2015.

5.2 Monte Carlo analysis

We also undertook a Monte Carlo simulation analysis using the data on the twenty years of returns of the active and passive portfolios (shown in Table 3). In each iteration of this simulation, a fixed number of years' (5, 10, 15, or 20) returns were randomly drawn and the average calculated for that set of returns for each portfolio. For example, if a set of five random returns were drawn, those returns were not necessarily for five consecutive years, but were any random five years within the twenty-year period, 1995 through 2015. The simulation was undertaken with 50,000 iterations for each set of years (5, 10, 15, or 20). The results of the simulation analysis are shown in the Table 4.

Not surprisingly, the average returns of the two portfolios, for the various sets of years, are very close to the single-pass average returns shown in Table 3. Table 4 also reports the percent of the 50,000 random draws in which the active portfolio's average return is better than the passive portfolio's. These results suggest that the outperformance of the active portfolio is not driven by a set of superior returns, which are clustered in the recent or in the early years of the twenty-year period. Furthermore, because the years in each iteration are randomly selected, the simulation allows for examination of the two portfolios' performance in different market environments. The results of the Monte Carlo simulation provide additional support for the robustness of the paper's key finding.

5.3 Portfolio turnover and transaction costs

To the extent that the active and passive portfolios are reconstituted each year by selecting the largest funds in each asset class in each category, the investor would incur transaction costs in the form of trading costs and, potentially, capital gains taxes. However, it is unlikely these costs would alter our findings, and, in fact, may even add to the performance advantage of the active portfolio.

First, these costs would increase with the frequency of the portfolio turnover, i.e., reconstitution of the portfolio constituents. However, the data show that this reconstitution is infrequent for either portfolio. This is because the largest funds, in both management styles, have a high degree of persistence: the largest fund in a category in any given year continues to be the largest fund in subsequent years. Specifically, out of the 15 largest funds in the active and passive portfolios, typically only one or two funds change per year, over the 20 years considered. Both portfolios also have 13 funds that are held for more than 10 years out of the 20-year period.

Second, we evaluated the taxable gains and losses from each portfolio based upon the sales necessary to rebalance and reconstitute the portfolio each year. Overall, we found that the passive portfolio has larger taxable gains regardless of how one accounts for carry-forward losses. As a result, the impact of capital gains taxes does not alter our key finding that the active portfolio outperforms the passive portfolio. Thus, it is highly unlikely that transaction costs would change the results of the horse-race.

5.4 Characteristics of the largest active funds

Recall that the results in Table 2 showed that on average actively managed funds did not outperform their passive counterparts in two of the three asset categories, and a majority of fund assets reside in these two categories. The horse race results in Table 3 paint a different picture. The explanation for this difference must be that the largest active funds have different characteristics than the rest.

Table 4: Results of the Monte Carlo simulation

NO. OF YEARS	AVERAGE RETURN		DIFFERENCE	% OF DRAWS ACTIVE BETTER
	ACTIVE	PASSIVE		
5	5.85%	5.54%	0.312%	59.5%
10	5.86%	5.55%	0.306%	64.4%
15	5.88%	5.57%	0.311%	68.1%
20	5.85%	5.54%	0.311%	71.2%

Table 5: Largest active versus other active funds (1996-2015)

	U.S. EQUITY		FIXED INCOME		INTERNATIONAL EQUITY	
	LARGEST	OTHERS	LARGEST	OTHERS	LARGEST	OTHERS
Number	11	3,549	16	2,356	11	1,212
Average annual return (%)	7.2%	6.0%	4.5%	4.0%	6.1%	4.5%
Standard deviation	17.8%	19.2%	3.7%	4.7%	21.1%	23.6%
Average Sharpe	0.27	0.23	0.58	0.38	0.18	0.11
Average expense ratio (%)	0.91%	1.47%	0.69%	1.05%	1.16%	1.72%

We explore this issue by examining a related but distinct issue: the persistence of size for these active funds. That is, how frequently do the same funds appear in the list of the five largest funds across the years? If every year a new set of active funds were the largest, then, over the twenty years, one would have observed 100 unique funds in each asset category (20 times 5 funds/year); conversely, if there were no churn at all, then the same five largest active funds would have persisted across all 20 years. Consequently, in each asset category, the range for the number of unique active funds is bounded by a floor of 5 and a ceiling of 100. Table 4 shows that the largest active funds display a fair degree of persistence of size; the number of unique funds in the U.S. equity, fixed income, and international equity categories are: 11, 16, and 11, which means that the numbers of the unique funds were approximately a tenth of the maximum possible number of funds.

Table 5 presents a comparison of the performance characteristics of these largest active funds and other active funds in each asset category. The average for each metric is computed over the years of a fund's existence within the sample period, 1996-2015.

Table 5 shows that the largest funds have performed markedly better than the remainder of the active funds. In all three asset categories, the largest active funds have demonstrably better absolute and risk-adjusted performance and substantially lower fees than the other active funds. Given these characteristics, it is not surprising that investors have shown a preference for these funds and they have continued to grow over time, explaining the phenomenon of size persistence.

Our finding regarding the outperformance of the larger funds is consistent with the prior academic literature. A key explanation for this characteristic of the largest mutual funds is increased efficiencies due to economies of scale. Some examined benefits of economies of scale generally include: greater ability to diversify risk through larger capital pools and more efficient use of information, as well as a reduction of labor, risk absorption, and physical capital costs [Hughes and Mester (1998), Bossone and Lee (2004)].

Studies have also confirmed that economies of scale play a role in the mutual fund industry. Banko et al. (2010) write that "The economies of scale have the effect of decreasing fund expenses and hence cost to investors… In summary, our results suggest that a mutual-fund investor should invest in a fund, matching the investor's individual goals, that is of sufficient size to have significant scale economies" (p. 335). Other studies have also found that due to fixed costs of investing, there exists significant economies of scale in the mutual fund industry [Latzko (1999)].

Another reason presented for the better performance of these largest mutual funds is the impact of increased (decreased) fund flow due to higher (lower) performance. Matallin-Saez (2011) looked at Spanish mutual funds and determined that the outperformance of larger funds was due to increasing fund flows rather than the initial size of the fund itself. He argued that his analysis implied that, "a positive relation between average size and performance would not be due to the causality initially supposed from the economies of scale hypothesis; in other words, in any case, the performance would have caused an increase in fund size, and not the other way [a]round."

6. IMPLICATIONS

This paper showed that over the twenty-year period, 1996-2015, the average net of fees performance of actively managed U.S. equity and of fixed income funds was worse than that of passive index funds in these two categories. For international equity funds, however, actively managed funds outperformed the passive index funds. These results are generally consistent with the existing literature.

The result unique to this paper is: an investor would be better off by following a readily-implementable strategy of investing in a portfolio of the five largest active funds in each of the three asset categories than investing in a corresponding portfolio of passive funds. It is important to note that this paper's methodology is free from hindsight bias because the portfolio was reconstructed every year based on which funds were the largest ones in the preceding year. This paper shows that the portfolio of the largest actively managed funds outperformed a similarly constructed portfolio of passive funds.

The active-fund-portfolio outperformed not only in terms of average returns, but also in risk-adjusted returns, providing far greater downside risk protection than the passive fund portfolio. These findings call into question the veracity of the "wisdom" of index investing, which has been receiving a lot of attention in the financial press in the recent years. Furthermore, the results have important implications for financial advisors' mutual fund recommendations, particularly in light of the U.S. Department of Labor's new fiduciary rule, which will govern the way advisors help their customers invest for retirement.

Given the findings of this study, investors and their advisors should consider the potential benefits offered by a portfolio of the largest actively managed funds. While the popular press and new guidelines might suggest that passive funds offer superior returns, this analysis clearly shows that active funds can provide superior returns when compared with a portfolio of similar passive funds.

References

Anderson, T., 2015, "Index funds trounce actively managed funds: study," CNBC, June 26, http://cnb.cx/2uahyN5

Banko, J., S. Beyer, and R. Dowen, 2010, "Economies of scope and scale in mutual-fund industry," Managerial Finance 36:4, 322-336

Bossone, B., and T-K. Lee, 2004, "In finance, size matters: the 'systemic scale economies' hypothesis," IMF Staff Papers 51:1, 19-46

Carhart, M. M., 1997, "On persistence in mutual fund performance," Journal of Finance 52:1, 57-82

Clements, J., 2015, "Are index-fund investors smarter?" Wall Street Journal, March 26, http://on.wsj.com/2vWGZnw

Fama, E. F., and K. R. French, 2010, "Luck versus skill in the cross-section of mutual fund returns," Journal of Finance 65:5, 1915-1947

Fama, E. F., K. R. French, D. G. Booth, and R. Sinquefield, 1993, "Differences in the risks and returns of NYSE and NASD stocks" Financial Analysts Journal 49:1, 37-41

Grinblatt, M., and S. Titman, 1989, "Mutual fund performance: an analysis of quarterly portfolio holdings," Journal of Business 62:4, 393- 416

Grinblatt, M., and S. Titman, 1993, "Performance measurement without benchmarks: an examination of mutual fund returns," Journal of Business 66:1, 47- 68

Holmes, M., 2007, "Improved study finds index management usually outperforms active management," Journal of Financial Planning 20:1, 48-58

Hughes, J. P., and L. J. Mester, 1998, "Bank capitalization and cost: evidence of scale economies in risk management and signaling," Review of Economics and Statistics 80:2, 314-325

Jensen, M. C., 1968, "The performance of mutual funds in the period 1945-1964," Journal of Finance 23:2, 389-416

Johnson, B., T. Boccellari, A. Bryan, and M. Rawson, 2015, "Morningstar's active/passive barometer: a new yardstick for an old debate," Morningstar, June, http://bit.ly/1rleSLw

Kacperczyk, M., C. Sialm, and L. Zheng, 2005, "On the industry concentration of actively managed equity mutual funds," Journal of Finance 60:4, 1983-2011

Krouse, S., and C. Driebusch, 2016, "Investors snub money managers for market clones," Wall Street Journal, January 13, http://on.wsj.com/1mYSJAL

Lamy, A., and T. Strauts, 2015, "2014 global asset flows report," Morningstar, March 24, http://bit.ly/2uOCURT

Latzko, D. A., 1999, "Economies of scale in mutual fund administration," Journal of Financial Research 22:3, 331-339

Malkiel, B. G., 1995, "Returns from investing in equity mutual funds 1971 To 1991," Journal of Finance 50:2, 549-572

Matallín-Sáez, J. C., 2011, "On causality in the size-efficiency relationship: the effect of investor cash flows on the mutual fund industry," Applied Economics 43:27, 4069-4079

Rich F., and S. Michelson, 1999, "Fund indexing vs. active management: the results are..." Journal of Financial Planning 12:2, 74-81

Rich F., and S. Michelson, 2002, "Indexing versus active mutual fund management," Journal of Financial Planning 15:9, 82-94

Sirri, E. R., and P. Tufano, 1998, "Costly search and mutual fund flows," Journal of Finance 53:5, 1589-1622

Sisti, G., 2014, "Index funds superior to high-priced active management," Market Watch, August 1, http://on.mktw.net/2uOlxQb

Wermers, R., 2000, "Mutual fund performance: an empirical decomposition into stock-picking talent, style, transactions costs, and expenses," Journal of Finance 55:4, 1655-1695

Zoll, A., 2014, "A bull market in passive investing," Morningstar, January 6, http://bit.ly/2tPSvPY

Aligning interests over the long term: An incentive structure for U.S. 501(c)(3) private foundations

CHRISTOPHER RAPCEWICZ[1] | Director of Investment Risk Management and Operations,
The Leona M. and Harry B. Helmsley Charitable Trust

ABSTRACT

The Great Financial Crisis (GFC) demonstrated the consequences of adverse incentives. Although the principle that the risk to the agent's incentive payout should be as similar as possible to the risk experienced by the principal is easy to state, it is very difficult to achieve in practice. An incentive structure for a U.S. foundation that must make a return on its endowment to offset its required 5% annual payout is proposed. The incentive structure is a combination of call spreads linked to the performance of the foundation's invested assets and put spreads linked to the performance of the foundation's invested assets whose payout is the carry-forward of investment staff's unearned incentive compensation. With this structure, the risk of the staff's incentive compensation is similar to the investment risk of the foundation's invested assets. One of its benefits is that it further encourages investment staff to protect capital when necessary.

[1] This research is that of the author alone and does not reflect the views of the Helmsley Charitable Trust

1. INTRODUCTION

Every organization faces the challenge of aligning the activity of its staff with its goals. This situation arises because the staff acting as an agent of the firm (or so-called principal) might have different goals. Consequently, the principal is well-advised to try to ensure that the interests of both parties are aligned. Depending upon the principal's goals, it may not be possible to align interests; consequently, the principal may have to change goals in order to achieve alignment. Agency Theory [Gibbons (1998), Eisenhardt (1989), Roberts (2011)] attempts to address this relationship. Historically, it has concentrated on resolving two problems that can occur in the agency relationship: the misalignment of interests of the organization and the staff, and the verification of the staff's activities.

Experience has shown that aligned goals can be very difficult to achieve. Kerr (1975) presents a long list of examples that highlight an additional risk that a principal faces when attempting to align goals, namely that the employee incentives actually misalign interests. Indeed, in the wake of the GFC, the Financial Stability Board (FSB) published its "Principles for sound compensation practices" [FSB (2009)], which identified the incentive structures at financial institutions as a key causal factor in the GFC, because they created "perverse incentives [that] amplified [...] excessive risk-taking." The "Principles" called for the financial services industry to align employees' goals with the long-term profitability of the company. Post GFC, reviewing the theory and evidence on incentive pay, Roberts (2011) commented that "it is unwise to give strong incentives based on only some aspects of the overall risks and returns that the Agent's actions generate." Indeed, the GFC made very clear that the key problem with aligning incentives is that it is not at all obvious how to create an incentive structure in which the risks to the agent's incentive payout mirror sufficiently closely the principal's investment risk.

In this article, we consider the case of a U.S. foundation that has been mandated to survive into perpetuity. It faces the challenge of how to achieve sufficient returns over time to offset the required 5% annual payout, the loss of purchasing power due to inflation, and any costs associated with managing the foundation's assets. For this reason, foundations hire investment staff to manage the investible assets of the foundation and set the goal of ensuring that the it survives into perpetuity. The foundation also uses performance-linked financial incentives to align the pay of the investment staff with the achievement of the foundation's objectives.

We develop an incentive structure for the investment staff of a U.S. 501(c)(3) private foundation that invests its endowment. The risk of the staff's incentive payout is designed to be similar to the investment risk of the foundation's invested assets. This incentive structure cannot be achieved with a simple combination of options on the performance of the foundation's endowment.

2. THE FOUNDATION'S AND STAFF'S PREFERENCES

It is worthwhile to consider in detail the foundation's reasons for offering financial incentives and the staff's preferences for an incentive structure. Survival of the foundation is the sine qua non for the foundation to ensure its charitable objectives. The foundation provides competitive (including incentive) compensation in order to attract and retain the high-caliber talent needed to manage the assets. The foundation, however, is not indifferent to how the staff achieve returns on the investment assets. A foundation, like most organizations, prefers a more stable return stream to one in which the value of the assets can fall significantly and unpredictably, resulting in the foundation terminating grant-making staff, defaulting on grants, or having the endowment impaired permanently. Further, the foundation prefers that the incentive structure keep the staff continuously focused on achieving the objective.

The staff's preferences for incentive compensation are well aligned with the foundation's desired outcomes. The staff prefer that the full incentive compensation should be received if the foundation's investment objective is achieved. The staff, too, prefer a stable salary to a very volatile one. They are also aware of the trade-offs between working in a commercial organization and a foundation and, therefore, should be less inclined to sacrifice the foundation's interest to their own interest. The staff, like all market participants, are forward looking. Their preference for more control over their salary means that they prefer that the incentive compensation not be structured around a single instant at which the payout or some portion of it is received and the rest lost but have a more continuous character; this aligns very well with the foundation's desire for a staff continuously focused on achieving the foundation's goals.

Although the foundation's and staff's preferences for incentive compensation are initially well aligned, the structure of the incentive compensation could misalign the goals of the foundation and staff. As Ordóñez et al. (2009) note, "goal setting is one of the most replicated and influential paradigms in the management literature" that can "inspire employees and improve performance" but "there are many ways in which goals go wild: they can narrow focus, motivate risk-taking, lure people into unethical behavior, inhibit learning, increase competition, and decrease intrinsic motivation." The authors provide a list of ten questions to consider when developing an incentive structure:

1. Are the goals too specific?
2. Are the goals too challenging?
3. Who sets the goals?
4. Is the time horizon appropriate?
5. How might goals influence risk taking?
6. How might goals motivate unethical behavior?
7. Can goals be idiosyncratically tailored for individual abilities and circumstances while preserving fairness?
8. How will goals influence organizational culture?
9. Are individuals intrinsically motivated?
10. Consider the ultimate goals of the organization and what type of goal (performance or learning) is most appropriate?

Two common types of incentive structures used in finance are the so-called "trader's option" and the benchmarked relative-performance incentive structure. The following analysis demonstrates that both of these incentive structures are perverse and identifies their flaws. An incentive structure that does not have these problems is presented.

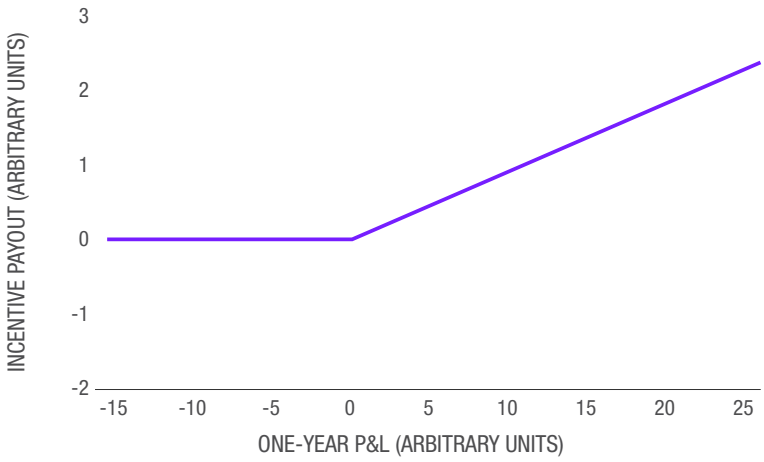
3. THE TRADER'S OPTION

The trader's option is an incentive structure that was originally employed by banks to pay traders on the trading floor. It is very simple: if, at the end of the year, a trader has made a profit, they receive a percentage of that performance; if they made a loss, then they would not receive any payout. Figure 1 shows the trader's incentive payout at year end as a function of profit and loss (P&L). It has the form of a long call-option on the P&L in which the strike price is zero P&L.

The incentive structure in option terms provides insights into how to improve the structure's design. The traders hold a long call-option on their P&L or, equivalently, the bank is short an option. This simple option can be priced with the Black-Scholes formula from which it is clear that the traders can increase the value of their long option by increasing the volatility of their P&L, i.e., by taking more risk. For this reason, the bank is said to be short volatility and the trader long volatility (volatility being a proxy for risk in the option pricing formula). The traders are incentivized to take as much risk as possible: if successful, they receive a handsome payout; if they suffer large losses, they lose their jobs. In the latter case, the bank is responsible for the losses. Taleb (2008) provides a useful analysis of this behavior. This asymmetry in the incentive structure is called the trader's option. The asymmetric nature of the payoff can encourage staff to take excessive risk (from the bank's and society's perspective) in order to get a larger incentive payout.

The single point-in-time nature of the incentive compensation with its bullet payout can also distort the behavior of the trader. The trader can follow a strategy to take advantage of the single point-in-time nature of the payout: they can create complex financial instruments in which there is a likelihood of underperformance in some distant future. In this way, the trader could enjoy several years of payouts and leave the bank before the instruments blow up. This is exactly what happened during the GFC. In good times, traders at the banks created liabilities for the banks that were conditional on events that were thought to be unlikely and, therefore,

Figure 1: The trader's option



low risk. The traders received their incentive payouts in the years that the bank was paid to assume these conditional liabilities. The actual risks were realized only later during the GFC. The so-called CDO liquidity puts were an example of this phenomenon [Jones (2007)].

Another way in which the structure of the compensation can impact an organization is through its signaling effect. Risk-seeking individuals would be very attracted to the possibility of working for an organization that offered a trader's option and would pursue such employment opportunities aggressively. Gordon Gekko [Stone (1987)], the Masters of the Universe and Big Swinging Dicks [Lewis (1989)] demonstrate that the incentive structure was so successful in attracting a certain type of risk-seeking individual that the personality type has become a staple of popular culture. As a result, the traders at the banks could become disproportionately composed of aggressive risk-takers, which would result in a higher likelihood of a blow-up or multiple blow-ups, situations in which traders took excessive risks and the bank suffered heavy losses [Taleb (2008)].

Banks realized the problems associated with this asymmetric payoff and tried to "hedge" their short volatility by introducing risk management into the organization. Risk managers received a relatively constant salary and were held responsible if a trader whose risk they were "managing" were to suffer a large loss. Thus, the situation of the risk manager was meant to be specular to that of the trader. But, there was a mismatch, a basis risk; while the trader was long a call option on his/her P&L, the risk manager lost his/her job instead of having to repay the losses; and the bank still bore the financial responsibility. Of course, the risk manager's incentive structure had its own limitations [Taleb (2008)].

After the GFC showed the perversity of the single point-in-time bullet payments, banks introduced deferments and claw-backs in an attempt to mitigate the time-horizon risk embedded in their flawed incentive structure. From the perspective of "goals gone wild," the trader's option created very different risk profiles for the bank and the trader, with the consequence of a culture heavy with aggressive risk-takers and excessive risk-taking. The time-horizon of the incentive structure narrowed the focus to the current year and further misaligned interests.

It is not clear whether the incentive structure was inconsistent with the bank's ultimate goals since the reason for offering the unlimited upside of the trader's option to its trading staff was the bank's desire to maximize profits.[2] High leverage and complex instruments used by banks exacerbated the consequences of the perverse incentive structure.

Foundations are not highly leveraged and they do not make use of complex, levered, financial structures to take risk. Nonetheless, an incentive structure that does not cause the foundation to be short the trader's option is desirable.

"Unless the market risk of the staff's incentive payout and the foundation's invested assets align, the incentive structure will be adverse; consequently the performance reference for any incentive structure must be the performance of the foundation's investment assets."

4. CHARACTERISTICS OF INCENTIVE STRUCTURES

The previous example shows how a well-meaning, but poorly designed, incentive structure can result in unforeseen, negative outcomes. It makes sense to analyze any proposed incentive structure in detail in order to ascertain what risks each aspect of the incentive structure poses and how they can be combined in order to align the investment staff with the foundation. The incentive structure consists of four characteristics: the base salary, the payout formula, time dependence of the pay off, and reference for the payout. An incentive structure is most effective if it aligns the goals of the staff and the foundation.

A poor choice of any of the characteristics will result in a perverse incentive structure that can, in theory, fail to align interests and leave the foundation vulnerable to principal-agent problems. For example, if the foundation sets the performance-linked compensation too small, the staff may be indifferent to the investment outcome; if the foundation sets the base compensation too low, the foundation may risk high turnover of staff and excessive risk-taking to achieve the payout; if the foundation makes it too difficult to achieve the

[2] There is one institution that has managed, so far, to make this perverse incentive structure work. The organization is a hedge fund, is owned by a single person and, hence, has one principal, employs rigorous risk management, traders are fired, almost without exception, if they suffer a draw-down greater than a set amount, and, the hedge fund trades only liquid market-traded securities.

incentive compensation, the staff may act as if it does not exist and be indifferent to the investment outcome; if the foundation links the achievement of the incentive compensation to the return during a single year, staff's attention may be myopically focused on the current-year's return; if the foundation judges the performance of staff relative to a reference that is unrelated to the desired performance of the foundation's assets, staff may be focused on a task that, in the best case, will be peripherally related to the foundation's objectives and, at worst, unrelated. Notwithstanding any misalignment that may occur as a result of a flawed incentive structure, the good will of the staff and their desire for success of the foundation's mission may mitigate the problems of the incentive structure.

5. BENCHMARKED RELATIVE-RETURN INCENTIVE STRUCTURE

Foundations often substitute the return of a market-based, passive benchmark for the performance of the foundation's investment portfolio when determining the incentive compensation for the staff. This passive benchmark normally consists of an allocation to equity, fixed income, and, possibly, alternative indices. The staff receive their incentive compensation based upon outperformance of the market benchmark, up to some maximum. Often, the outperformance is calculated on a rolling three-year basis, paid out once a year. Does this incentive structure help the foundation achieve its desired outcome?

Measuring performance relative to the three-year cumulative return of the benchmark together with the single point-in-time nature of the performance-linked payout may better condition the behavior of the staff. Staff is motivated to deviate from the benchmark in a way that increases the probability of achieving the incentive payout. If the staff do decide to deviate, it will be to take more risk rather than less risk. This is because it is generally believed that the equity markets go up, so that taking more risk than the benchmark increases the likelihood outperforming it. As a result, the foundation's portfolio risk should be expected to be higher than the benchmark. This tendency to take unnecessary risk will be further exacerbated if peer performance plays a role in the evaluation of the staff. If one takes more risk than one's peers, the reasoning goes, one is more likely to outperform them (except in those unfortunate years in which there is a market sell-off). Indeed, encouraging staff to achieve a high peer ranking is equivalent to encouraging the staff to take excessive risk.

The benchmarked relative-return incentive acts as a perverse incentive-structure that creates a situation in which the staff's risk is misaligned with the foundation's risk. For the foundation, reducing risk means reducing absolute risk, i.e., increasing the allocation to cash and short-term U.S. Treasurys. For the staff, reducing risk means moving allocations closer to passive, benchmark allocations, which could in certain circumstances result in an absolute increase in risk. This is misalignment. The point-in-time aspect of this incentive structure misaligns staff's interest from the foundation's by myopically focusing the staff on the current year's performance and discouraging them from acting to preserve capital. In practice, the impact of the misalignment may be ameliorated due to the type of person that is attracted to working at a foundation.

Finally, there is no certainty that a passive, market benchmark will achieve the performance goals of the foundation. Consequently, the staff could end up with their incentive compensation determined by a performance goal unrelated to the return objective of the foundation.

From the perspective of "goals gone wild," the benchmarked relative-return incentive structure creates overly specific goals and influences risk taking by decoupling the staff's outcomes from that of the foundation's. The single point-in-time payout focuses the staff's attention on the current year instead of a longer horizon.

6. CALL-SPREAD PUT-SPREAD (CSPS) INCENTIVE STRUCTURE

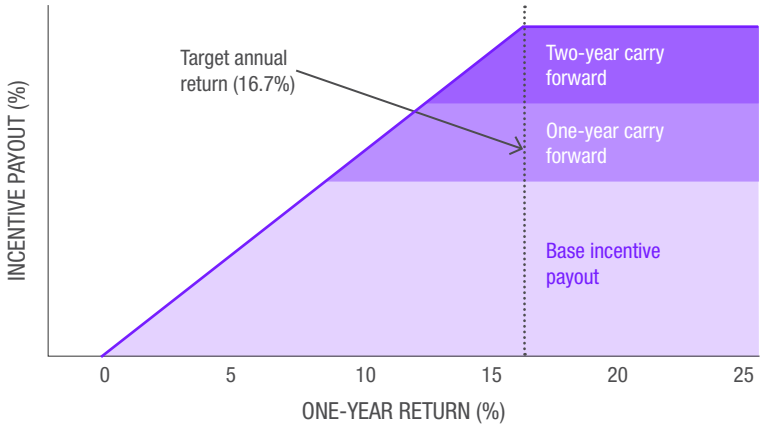
Is it possible to design a better incentive structure? The discussion above showed the two key factors that result in a perverse incentive structure: a misalignment of the risks between the staff and the organization and the bullet-style point-in-time nature of the incentive compensation structure.

The analysis of the benchmarked relative-performance incentive structure showed that unless the staff's and the foundation's risks align, the incentive structure will be perverse; consequently the reference for any incentive structure must be the performance of the foundation's investment assets. The analysis of the trader's option revealed the short-option nature of the incentive structure that banks tried to "hedge" by introducing risk management. The most effective way to deal with the short-volatility exposure that led to excessive risk taking is to offset it with a long-volatility exposure. Instead of trying to "hedge" organizationally, the banks should have created an incentive structure that naturally hedged their short volatility by embedding short optionality in the trader's payoff, which would have discouraged excessive risk-taking; they would have better mitigated the risk associated with the trader's option. The structure that offsets the short-volatility with a long-volatility exposures is a call spread linked to the performance of foundation's investment assets.

The analyses above showed bullet-style payouts result in flawed incentive structures. The benchmarked relative-performance incentive structure tried to mitigate this risk by using a three-year rolling performance as the reference. The incentive structure is perverse because the staff are incentivized to take excessive risks and not to preserve capital when necessary, as de-risking the portfolio results in the staff taking significant risk. This can be resolved by changing the structure from a bullet-style payout to multi-year payout in which unattained incentive compensation can be carried forward for a period of time. This encourages the staff to protect capital by rolling forward the incentive payout to the future when economic and market conditions have improved. The payout structure for a single year is shown in Figure 2.

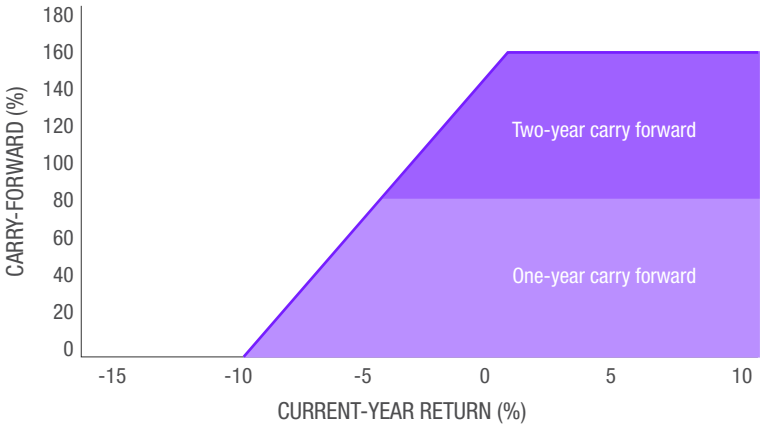
The introduction of an explicit target return provides the foundation with a new tool that can be used to align

Figure 2: The call-spread payout



The strategic return-objective is assumed to be 7%, the previous year's performance at 5%, and the return of two year's prior at 0%. The one-year and two-year carry-forwards depend on the returns of the two prior years

Figure 3: The put-spread embedded in the unearned incentive carry-forward



The upper strike is 0% and the lower strike is -10%. If the fund were down by 5% then only the unearned payout from the current year would be carried-forward; the unearned payout from the prior year would be reduced to zero because the staff did not protect capital.

goals, namely the portion of the incentive compensation that is not earned in the current year. Providing staff with the possibility of achieving the unearned portion of a previous year's incentive compensation obviously has positive value. The foundation can make its attainment contingent upon the staff having made up the previous year's underperformance and not having suffered a draw-down. This is achieved by embedding a short put spread in the unearned incentive-payout carry-forward. The put spread makes the staff more sensitive to the impact of draw-downs and signals the maximum draw-down that the foundation is willing to undergo in order achieve its target return. If the draw-down is greater than a certain amount, there is no carry-forward of the unearned incentive payout. The upper strike of the put spread determines the loss at which the carry-forward payout begins to decrease and the lower strike determines the loss at which no unearned incentive payout is carried forward to the next year. This carry-forward is shown in Figure 3 for the case that two years of unearned incentive payouts are being carried forward.

The choice of the number of years over which to evaluate the performance is important. The longer the period, the more tightly aligned are the interests of the foundation and the staff. However, a problem with too long of a period is that at normal levels of employee turnover, the staff could find themselves in a situation where the current staff was not responsible for the bulk of the performance. Any finite look-back invariably introduces a basis effect when large positive or negative returns are no longer included. A three-year look-back is a common compromise so that unearned compensation would carry-forward for two years (the n-year look-back should always be paired with an n-year period to achieve the unearned payout in order to align the interests, which means a carry-forward for n-1 years). In the rest of the article, a three-year look-back will be assumed for convenience.

The target return for a given year is the performance needed to ensure that the three-year cumulative performance of the foundation's asset is equal to the three-year target performance.

Table 1: Numerical example of the incentive structure for a typical case over 10 years with a three-year lookback

YEAR	1	2	3	4	5	6	7	8	9	10
3-year target	7.0%	14.5%	22.5%	22.5%	22.5%	22.5%	22.5%	22.5%	22.5%	22.5%
Target return r_{target}	7.0%	2.2%	0.3%	5.0%	10.1%	47.2%	33.2%	1.7%	4.5%	1.3%
Performance r_0	12.0%	9.0%	7.0%	4.0%	-20.0%	15.0%	4.7%	12.0%	8.0%	11.5%
Base incentive compensation I_{base}	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%	80.0%
First carry forward C_1	0.0%	0.0%	0.0%	0.0%	16.5%	0.0%	54.6%	60.9%	0.0%	0.0%
Second carry forward C_2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	54.6%	0.0%	0.0%
Target incentive compensation I_{target}	80.0%	80.0%	80.0%	80.0%	96.5%	80.0%	134.6%	195.5%	80.0%	80.0%
Pay-out percentage	1.00	1.00	1.00	0.79	0.00	0.32	0.14	1.00	1.00	1.00
Achieved incentive compensation $I_{achieved}$	80.0%	80.0%	80.0%	63.5%	0.0%	25.4%	19.1%	195.5%	80.0%	80.0%
Unearned base incentive compensation C_0	0.0%	0.0%	0.0%	16.5%	80.0%	54.6%	60.9%	0.0%	0.0%	0.0%
Unearned first carry forward C'_1	0.0%	0.0%	0.0%	0.0%	16.5%	0.0%	54.6%	0.0%	0.0%	0.0%
Haircut *h*	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
Unearned base incentive compensation after Haircut C^f_1	0.0%	0.0%	0.0%	16.5%	0.0%	54.6%	60.9%	0.0%	0.0%	0.0%
Unearned first carry forward after Haircut C^f_1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	54.6%	0.0%	0.0%	0.0%

A call spread on the one-year target performance of the foundation's investment assets, in which any unearned incentive compensation is rolled forward for up to two years, is an incentive structure that resolves the problems highlighted above. The lower strike of the call option should be set at zero and the upper strike to the annualized three-year target return. Any unearned incentive compensation from one year can be earned in the subsequent two years (for a total of three years) subject to being haircut as a result of negative performance. This incentive structure avoids both problems and is summarized in detail in the Appendix at the end of the article.

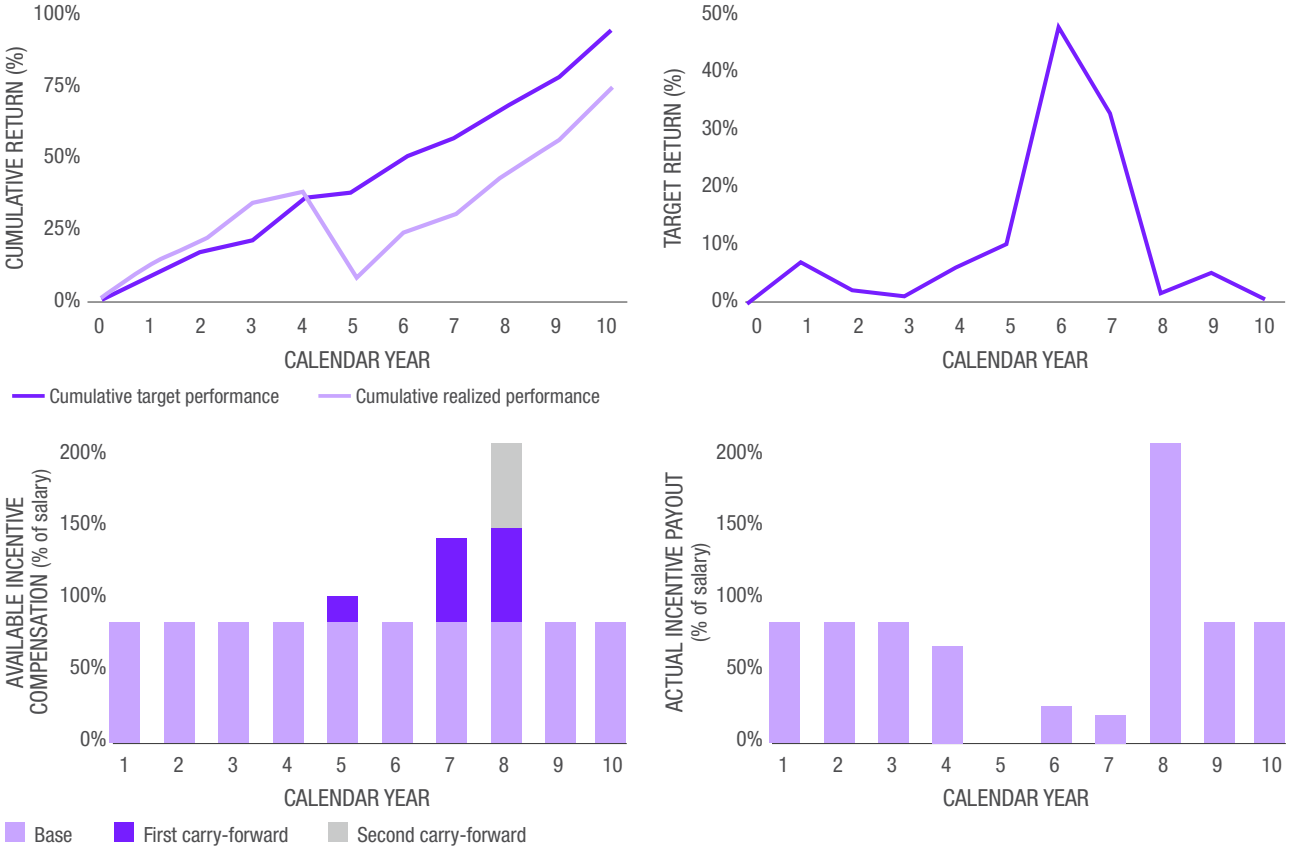
The incentive payout at the end of the year is equal to the available incentive payout times the ratio of the current year's return divided by the target return. Any unearned payout is carried forward for up to two years and is earned in order of decreasing priority: current year payout, then the unearned compensation from the previous year, and finally the unearned payout from two years ago.

7. NUMERICAL EXAMPLE

Table 1 provides a detailed numerical example of how the incentive payout would work in a typical case over a ten year period. In the example, the strategic return objective r_{sro} = 7%, the cumulative performance is measured for three years (n = 3), and the incentive compensation is 80% of the base salary (l_{base} = 80%). In the example, the payout is quoted as a percentage of the base salary. For instance, if the base salary were $300K and the staff were to receive incentive compensation of 25.4%, then the staff would receive incentive compensation of $76.2K. Any carry-forward begins to be haircut if the performance of the fund is negative (X_{upper} = 0) and there is no carry-forward if the fund is down more than 10% (X_{lower} = 10%). The incentive structure thus signals to the staff to attempt to make 7% per annum and to preserve capital.

The first two years show how a transition to the incentive structure could work: in the first year, the return target (r_{target}) would be 7% and in the second year

Figure 4: The CSPS incentive structure for a numerical example with a three-year lookback



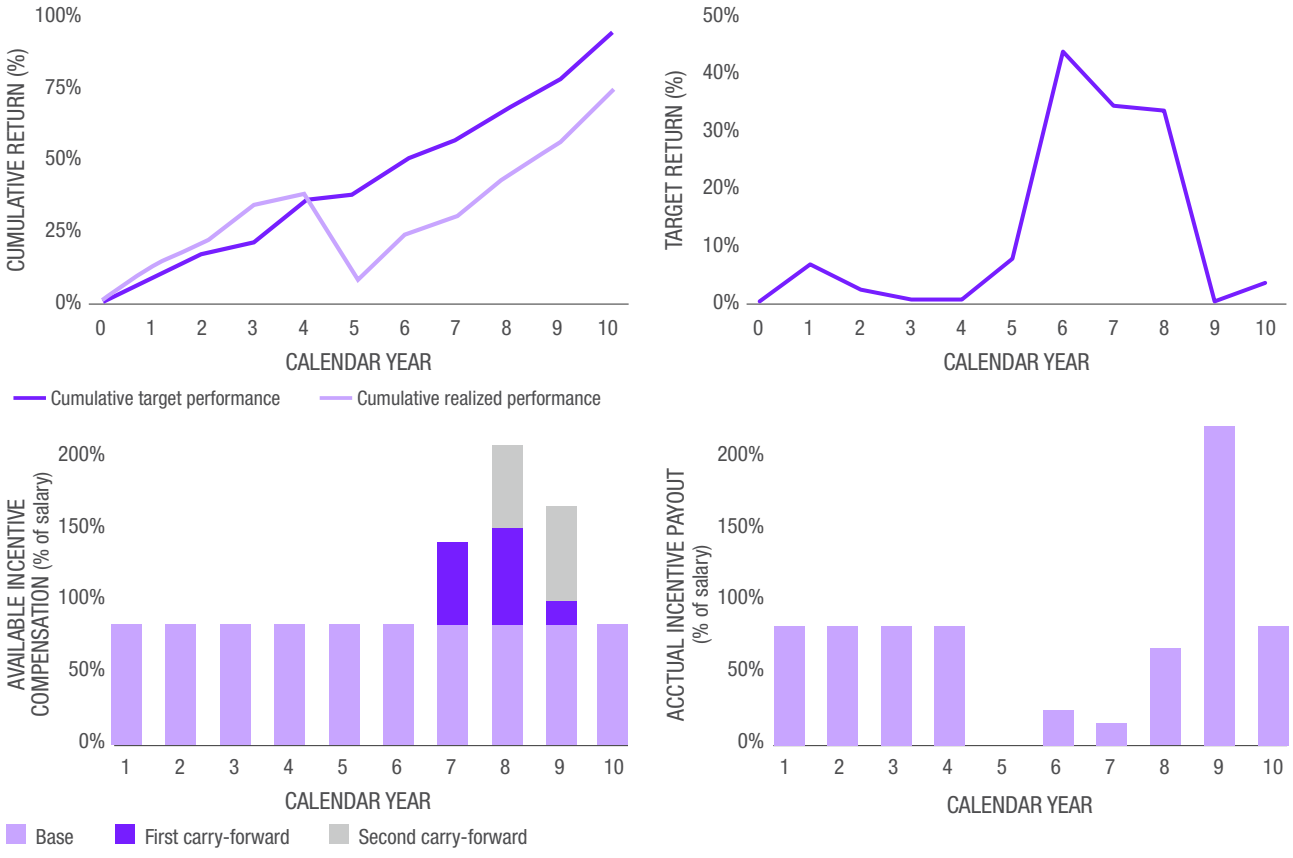
the target return would be calculated using only the one year return. By the third year, the incentive structure would be fully implemented.

In the hypothetical example, in year one, the one-year target return would have been 7%, the performance was 12.0%, and the staff would have received the entire incentive compensation (and there would not be any carry forward). In year 2, the cumulative return target would have been 14.5%. Since the previous year's return was 12.0%, the target return for the year would have been 2.2%. As performance was 9%, the staff would have received the entire incentive payout. In year 3, the three-year cumulative return would have been 22.5% (7% compounded for three years); since the cumulative return over the prior two years was 22.1%, the target return would have been 0.3%. The performance was 7% and the staff would have received the all of the performance compensation (80% of their base salary). Since the cumulative return of the prior two years was 16.6%, the target return in year 4 would have been 5%. As a result of the 4% performance, the staff would have received incentive compensation of 64%, with the remaining 16% carried forward to the subsequent year.

In year 5, the target bonus would have been 96% (of the base salary) but performance was -20%. Since the under-performance was less than -10%, the staff not only would not have received any incentive compensation, but there would be no carry-forward. As a result of one year's mediocre performance and the second year's poor performance, the target return for the year would have been 47.2% (the staff's cumulative two-year return was -16.8%) and the target bonus 80%. The 15.2% performance that year would have resulted in the staff receiving incentive compensation of 25%, leaving 55% to be carried forward to the next year.

In year 7, the target return for the year would have been 33.2% and the target bonus 135%. The 4.7% return for that year would have earned the staff incentive compensation of 19%, with 114% carried forward to next year. In year 8, the target return would have been 1.7% and the target bonus 196%. The 12% return would have earned the full 196% for the staff. In the

Figure 5: The CSPS incentive structure for the numerical example with a four-year lookback.



final two years, the staff's performance would have earned them the full bonus. Over the ten year period, the staff would have received 88% of the possible incentive compensation, in line with the fact that the NAV of the assets would have been 90% of the target NAV.

As noted above, increasing the look-back aligns more closely the interests of the staff and the foundation. Figure 5 shows for a four-year look back (n = 4).

8. FURTHER CONSIDERATIONS

The CSPS incentive structure is designed to encourage the staff to attempt to achieve the target return required by the foundation to survive into perpetuity while protecting capital. However, in the real world, the foundation may nonetheless suffer a significant draw-down. This was shown in the example. In the wake of the loss, it would be difficult, if not impossible, for the staff to achieve the entire bonus; however, they would be able to achieve some portion of the bonus that would mitigate the risk that they might leave in the face of an extended period without any bonus, as could happen if the staff underperformed under a "benchmarked relative-return incentive structure."

Although the CSPS incentive structure was designed for a foundation, it has wider applicability; it provides insight into the origin of the "perverse incentives [that] amplified [...] excessive risk-taking" [FSB (2009)] and played a significant role in causing the GFC. Indeed, the analysis presented here suggests that it may not be possible to align incentives if the principal demands that the agents maximize returns, since an unearned compensation cannot be defined due to the unlimited upside embedded in the incentive structure.

The CSPS also avoids a fundamental asymmetry between return and risk, namely that if one has two investments, A and B, and A has a higher expected return than B, then it is reasonable to assume that A has higher risk than B. However, if one has two investments, X and Y, and X has higher risk than Y, then one cannot assume that X has higher expected return than Y; one can easily find risky investments that have little or no expected return.

9. CONCLUSION

We presented an incentive structure that aligned well the interests of a U.S. foundation and its investment staff. An important prerequisite for the incentive structure was the recognition by the foundation that it was not trying to maximize returns each year but to achieve a specified target return over time. This permitted the creation of a new "asset," namely the current year's unearned incentive compensation. By carrying forward this unearned incentive compensation and giving the staff the opportunity to earn it contingent upon the portfolio having recovered from the underperformance

and not having suffered a draw-down deeper than a specified amount, the risk profile of the incentive compensation approaches closely the risk profile of the investment portfolio.

Wider implications of the insights gleaned when designing the incentive compensation were also noted.

APPENDIX: ALIGNING INTERESTS OVER THE LONG-TERM: STEP-BY-STEP

- Establish the one-year strategic return-objective, r_{sro};
- Establish the number of years, n, over which the performance of the foundation's assets will be computed (below three years will be assumed, so n=3);
- Establish the base incentive compensation payout I_{base};
- Establish the one-year negative return, X_{upper}, at which the carry-forward begins to be haircut (reduced);
- Establish the one-year negative return, X_{lower}, below which no unearned incentive compensation is carried forward;
- The incentive structure can be introduced by increasing each year the number of years over which the current-year's target performance is calculated: for a three-year look-back, only the current-year return is used in the first year and the incentive compensation is the base payout; in year two, the previous year's performance is used to calculate the target return and available incentive compensation is the base payout plus any carry-forward from the previous year. In year three, the incentive structure is fully implemented.

1. At the beginning of the year, the prior year's return r_1 and performance r_2 from two years before are already known and the target return can be determined:

$$r_{target} = \max\left[\frac{(1 + r_{sro})^3}{(1 + r_1)(1 + r_2)} - 1, 0\right] \quad (1)$$

2. The previous year's carry-forward C_1 is also known as is the carry-forward from two-years ago, C_2. The target incentive compensation, I_{target}, is:

$$I_{target} = I_{base} + C_1 + C_2; \quad (2)$$

3. At the end of the year when the performance, r_0, is known, the payout is:

$$I_{achieved} = I_{target} \times \min\left[\max\left(\frac{r_0}{r_{target}}, 0\right), 1\right]; \quad (3)$$

4. If the current year's performance was positive and the entire incentive compensation was not achieved, determine the carry-forward:

$$C_0 = \max(I_{base} - I_{achieved}, 0)$$
$$C'_1 = \max(C_1 - \max(I_{achieved} - I_{base}, 0), 0). \quad (4)$$

The total carry-forward is $C = C_0 + C^f_1$ and any unearned portion of C_2 is lost;

5. If the current year's performance was negative, determine how much the carry-forward should be haircut (recall $X_{lower} < X_{upper} < 0$):

$$h = C \times \min\left[\frac{(X_{upper} - \min(X_{upper}, r))}{(X_{upper} - X_{lower})}, 1\right] \quad (5)$$

6. If the carry-forward haircut is not zero, calculate the final, adjusted carry-forward:

$$C^f_1 = \max(C'_1 - h, 0)$$
$$C^f_0 = \max(C_0 - \max(h - C'_1, 0), 0) \quad (6)$$

7. The available incentive compensation for the next year is then:

$$I_{base} + C^f_0 + C^f_1 \quad (7)$$

References

Eisenhardt, K. M., 1989, "Agency theory: an assessment and review," Academy of Management Review 14:1, 57-74

FSB, 2009, Practices - Financial Stability Board, April

Gibbons, R., 1998, "Incentives in organizations," Journal of Economic Perspectives 12:4, 115–132

Jones, S., 2007, "The $25bn Citi CDO liquidity put – and who else has one," Financial Times, November 21

Kerr, S., 1975, "On the folly of rewarding A, while hoping for B," Academy of Management Journal 18:4, 769–783 Compensation

Lewis, M., 1989, Liar's poker, Penguin Books

Ordóñez, L. D., M. E. Schweitzer, A. D. Galinsky, and M. H. Bazerman, 2009, "Goals gone wild: the systematic side effects of overprescribing goal setting," Academy of Management Perspectives 23:1, 6–16

Roberts, J., 2011, "Weak incentives: when and why," in Marsden, D. (ed.), Employment in the lean years, Oxford University Press

Stone, O., 1987, "Wall Street," IMDb ID: tt0094291

Taleb, N. N., 2008, "Fooled by randomness: the hidden role of chance in life and in the markets, Random House

Financial inclusion and consumer payment choice

ALLISON COLE | Ph.D. Candidate, Massachusetts Institute of Technology[1]
CLAIRE GREENE | Payments Analyst, Consumer Payments Research Center, Federal Reserve Bank of Boston

ABSTRACT

This article examines similarities and differences among three groups of consumers: those without a checking or savings account (unbanked), bank account adopters who have used alternative financial services (AFS) in the past 12 months (underbanked), and bank account adopters who did not use AFS in the past 12 months (fully banked). Consumers in the three groups have different demographic characteristics, income, and payment behaviors. The payment behavior of the underbanked is similar to that of the fully banked; unbanked consumers make fewer payments per month than the fully banked and the underbanked; fewer than half of the unbanked know their credit scores, while about 85% of the underbanked and the fully banked know theirs; and both unbanked and underbanked consumers are significantly more likely than fully banked consumers to own a general purpose reloadable (GPR) prepaid card. We find no evidence that consumers are prevented from opening a bank account; many cite personal preferences and cost as reasons for choosing to be unbanked. These preferences are likely related to income constraints.

[1] This article is the revised version of the one available on the website of the Federal Reserve Bank of Boston at http://bit.ly/2w5vxoQ. Suzanne Lorant, Scott Schuh, Joanna Stavins, and Robert Triest provided helpful comments. The authors are responsible for any errors. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Boston or the Federal Reserve System.

1. INTRODUCTION

Many U.S. policymakers believe that access to safe and affordable financial services is important for dealing with unexpected expenses, avoiding unnecessary fees, establishing the ability to borrow, and saving for the future, and that lack of such access is a sign of financial and civic marginalization that public policy should address. In 2005, Congress mandated that the FDIC conduct surveys of banks' efforts to bring individuals and households into the formal banking system.[2] The FDIC notes that "public confidence in the banking system is strengthened when banks effectively serve the broadest possible set of consumers" [Burhouse et al. (2014)]. Moreover, the Council of Economic Advisors reports that lack of financial inclusion, in particular access to credit, has broad consequences for the macroeconomy [White House Council of Economic Advisers (2016)].

One aspect of financial inclusion is access to the mainstream payments system, which enables one to conveniently receive funds, make purchases, and pay bills. This article identifies consumers according to their banking status in order to see how they receive funds and make payments. We examine the demographic characteristics of three groups of U.S. consumers, classified according to their degree of attachment to the mainstream financial system,[3] as well as their assessment of payment instrument characteristics, adoption of nonbank payment accounts, and adoption and use of payment instruments. Understanding payment choices made by consumers – especially those with weak attachment to the banking system – is potentially useful for researchers and policymakers studying financial inclusion, for innovators designing new financial products, and for financial educators seeking to understand consumer decision making.

Data on the banking status of U.S. consumers are from the 2014 Survey of Consumer Payment Choice (SCPC), the seventh in a series of annual studies (2008–2016) conducted by the Federal Reserve Bank of Boston [Schuh and Stavins (2015a), Greene et al. (2016)]. This survey collects detailed information about the accounts consumers use to manage income and payments, including checking and savings accounts at traditional financial institutions as well as newer methods, such as PayPal, general purpose reloadable (GRP) prepaid cards, and payroll cards. It measures the adoption and use by consumers of nine common payment instruments, including the four payment instruments associated with a checking account [checks, debit cards, online banking bill payments (OBBP), and bank account number payments (BANP)] as well as cash. It asks consumers to assess various characteristics, such as convenience and cost, of the nine payment instruments.[4]

2. DEFINITIONS OF BANKING STATUS

Consumers can be classified into two groups: banked and unbanked. A banked consumer is an individual who has at least one checking account or one savings account at a bank, credit union, brokerage, or investment firm. An unbanked consumer has neither checking nor savings account and, therefore, limited access to the mainstream payments system because they cannot use payment instruments linked to a bank account.[5] An unbanked consumer could be unbanked by choice or because they have been denied a bank account for various reasons (insufficient ID, prior account closed with negative balance, for example).

Table 1: SCPC definitions of bank accounts

CHECKING ACCOUNT	SAVINGS ACCOUNT
An account that allows a customer to make payments or withdrawals as often as necessary, using checks, debit or ATM cards, online, or pre-authorized withdrawal payments. Some checking accounts pay interest on deposits and may be called money market checking accounts.	Savings accounts allow only a limited number of payments, withdrawals, or transfers. Savings accounts pay interest on deposits that is usually higher than the interest on interest-bearing checking accounts. Examples include traditional savings accounts, money market savings accounts, Christmas Club accounts, and Coverdell or 529 education accounts.

Source: Federal Reserve Bank of Boston

In the SCPC, individual consumers report how many checking and/or savings accounts they have at banks, credit unions, brokerages, or investment firms.[6] Consumers report all accounts held individually and also those held jointly with a spouse or partner. Accounts held individually by a spouse or partner or for business purposes are not included. (Table 1 shows the SCPC definitions of these accounts.) An unbanked consumer does not hold either of these types

[2] Federal Deposit Insurance Reform Conforming Amendments Act of 2005, Section 7.
[3] The classifications used in this article, which are defined and discussed in the next section, are those of the FDIC, and are used in the Survey of Consumer Payment Choice, which is the source of our findings and is discussed in the next paragraph.
[4] The results reported here include the 1,809 respondents from the RAND American Life Panel. See Greene et al. (2016) for details.
[5] In this article, we use the term "bank account" loosely to include a savings or checking account (including a money market checking account, to which some may refer simply as a "money market account") at a credit union, brokerage, or investment firm, as well as at a bank.
[6] The SCPC includes individual consumers in the noninstitutional population age 18 and above, rather than all consumers. It surveys individuals, not households.

of accounts. The SCPC asks unbanked consumers whether or not they have owned a bank account at any time in the past.

Banked consumers can be further divided into two groups: fully banked and underbanked. Unlike the definition of the unbanked, which is more straightforward, the definition of the underbanked is nuanced. Conceptually, the underbanked are a subset of the banked population who, for whatever reasons, are not fully served by mainstream institutions that offer depository services. These consumers go elsewhere for financial products and services of this type, despite having a bank account. Consumers who go outside the banking system for deposit and transaction-related financial services "may not receive the same level of safety and security provided by deposit insurance and various federal consumer protections that are guaranteed by law, ensured by supervision, and enforced through a system of ongoing examination," according to the FDIC (2014). It may be, however, that underbanked consumers receive other benefits from their choices.

To get at this concept of being underserved, the FDIC defines underbanked consumers as those with a bank account who have purchased any of five AFS – money order, cashier's checks, check cashing, remittances, and payday loans – from a nonbank (that is, not a federally insured bank or thrift) and/or who have used personal property to secure a loan at a pawn shop, used rent-to-own services, or taken out a tax refund anticipation loan within the preceding 12 months. Both banked and underbanked consumers have access to all the bank-account-linked payment instruments (paper checks, debit cards, bank account number payment (BANP), and online banking bill pay (OBBP)). Fully banked consumers do not use the AFS listed above.

In 2014, the SCPC added two questions to identify consumers who are "underbanked," aligning with the FDIC definition:[7]

1. In the past 12 months, did you use any services provided by a nonbank (such as the Post Office): money order or cashier's check, check cashing, remittance, payday loan?

2. In the past 12 months, did you use any other financial services: selling an item at a pawn shop, rent-to-own services, tax refund anticipation loan?

Future versions of the SCPC (2015 and later) disaggregate these two questions into eight yes/no questions so it is possible to identify consumers according to the particular AFS they used. This could assist in identifying consumers for whom use of AFS reflects lack of access or poor financial health versus those for whom use of AFS is a choice driven by temporary circumstances. For example, compare a consumer who takes out a payday loan with a consumer who purchases a money order. The need to take out a payday loan could be seen as an inability to deal with unforeseen expenses. It might signal a lack of a savings cushion for a financial emergency and/or inability to access less-expensive sources of credit, for example, credit card debt. In contrast, a consumer might buy a money order because a payee requires that form of payment, for example, for a deposit on the purchase of a used car. In this case, the choice to use the money order would be externally driven and not related to the consumer's financial situation, knowledge of financial products and services, or ability to access lower-cost payment instruments. These various motivations for using AFS make it difficult to understand whether or not underbanked consumers are truly underserved. A further refinement to the SCPC questionnaire would be to ask consumers how frequently they use the various AFS within a 12-month period. A consumer who rolls over payday loans from paycheck to paycheck, for example, is in a different financial situation from one who takes out one payday loan over the course of a year in order to avoid overdrawing for an emergency medical payment.

These considerations show the difficulties of defining the state of being underbanked. Other researchers take a broader view of financial access. The Center for Financial Services Innovation (CSFI) defines "financial health" as encompassing effective day-to-day financial management, ability to deal with unforeseen expenses, and ability to take advantage of opportunities leading to financial security and mobility [Gutman et al. (2015)]. Access to high-quality financial products and services is one aspect of the CSFI definition of financial health but quality is not necessarily associated with whether those services are provided by a bank, thrift, credit union, or by a nonbank, for example, Western Union or the U.S. Postal Service.

[7] The 2014 SCPC questionnaire omitted one financial product included in the FDIC definition: auto-title loans. According to the 2013 FDIC survey, auto-title loans contributed 0.3% to the results.

The Bank for International Settlements (BIS) uses the term "financial inclusion" to encompass the availability and use of financial services. This article does not address availability (geographic proximity, for example); it focuses on use and its prerequisite, ownership, or setup of the relevant financial tool. The BIS also looks at financial literacy and the availability of financing for small and medium-sized enterprises, two topics outside the scope of this article.

3. OWNERSHIP OF CHECKING AND SAVINGS ACCOUNTS

Consumer adoption of traditional financial institution accounts for checking and savings has been steady for decades. In 2014, the percentage of consumers who are banked was 91.7%, unchanged from 2013.[8] Consumer ownership of checking accounts was 90.7% and consumer ownership of savings accounts was 74.7%. Ownership of checking accounts has been steady since the SCPC began in 2008. Adoption of saving accounts declined in the years following the recession and has partially recovered since 2010.

It is difficult to ascertain the size of the unbanked population because these statistics are self-reported and unbanked consumers may be more difficult to reach than other consumers. In 2014, the World Bank estimated that 6% of U.S. adults were unbanked [Demirgüç-Kunt et al. (2015)]. In 2013 and 2015, the FDIC estimated that 7.7% and 7.0%, respectively, of U.S. households were unbanked [Burhouse et al. (2014, 2016)]. In 2014, the SCPC found that 8.3% of U.S. consumers were unbanked (calculated as 100% minus the percentage of consumers who owned a checking or savings account at a bank, credit union, brokerage, or investment firm in October 2014). From 2013 to 2014, the SCPC found no statistically significant change in the percentage of consumers identified as unbanked. There also was no statistically significant change in the percentage of consumers identified as unbanked from 2008 to 2014.

In 2014, about one-quarter of consumers with a bank account, or 22.3% of U.S. consumers, were underbanked, according to the SCPC.[9] Of these underbanked consumers, 91% had purchased any of the five services (money orders, cashier's checks, check cashing, remittances, and payday loans) from a nonbank and 26% had used personal property to secure a loan at a pawn shop, used rent-to-own services, or taken out a tax refund anticipation loan.[10] In 2014, 69.4% of U.S. consumers were fully banked.[11]

The underbanked consumers had shallower banking relationships. While, by definition, underbanked consumers have at least one bank account, they were less likely than fully banked consumers to have had either a checking account or a savings account (Table 2), and also less likely to have both. Of underbanked consumers, 71% had both checking and savings accounts compared with 83% of fully banked consumers, a statistically significant difference at the 5% level.

Table 2: Bank account ownership by banking status

	FULLY BANKED	UNDERBANKED
Have bank account (percentage)	100	100
Have checking account	99.7	95.4*
Have savings account	83.0	75.6*
Have both checking and savings	82.7	71.0*

Source: 2014 SCPC, Federal Reserve Bank of Boston [Greene et al. (2016)].
Note: * indicates a significant difference from the fully banked group at the 5% level.

[8] Unless otherwise noted, all data are weighted as described in Angrisani et al. (2016).
[9] In 2013, the FDIC found that 19.7% of households were underbanked and in 2015, the FDIC the figure was 19.9%. This difference is not statistically significantly different from the SCPC estimate, which measures consumers. The standard error of the SCPC estimate is 1.4%, for a 95% confidence interval from 19.5% to 25.1%.
[10] The percentage of all consumers who use these groups of services is not available due to questionnaire design. Auto title liens, an element of the FDIC definition, were omitted from the questionnaire but represent less than 1% of AFS products used, according to Burhouse et al. (2014). Future versions of the SCPC will ask both unbanked and banked consumers about their use of these services.
[11] Computed as all consumers minus banked consumers who used AFS and minus unbanked consumers.

Table 3: Demographic comparison, by banking status (percentage unless otherwise indicated)

	FULLY BANKED	UNDERBANKED	UNBANKED
Number	1362	334	85
Gender			
Male	47.0	52.1	49.7
Average Age (years)	49.6	45.3*	36.5*
Race			
White	83.5	64.5*	32.8*
Education			
No high school diploma	3.2	5.2	37.6*
Labor force status			
Unemployed and looking for work	4.0	6.4	33.4*
Marital Status			
Married	70.5	54.5*	29.2*
Household income			
Less than U.S.$25,000	14.8	31.3*	75.9*
Number of household members	1.2	1.5*	2.1*

Source: 2014 SCPC, Federal Reserve Bank of Boston. Note: * indicates significantly different from the "fully banked" group at the 5% level. Results are weighted.

4. DEMOGRAPHIC CHARACTERISTICS BY BANKING STATUS

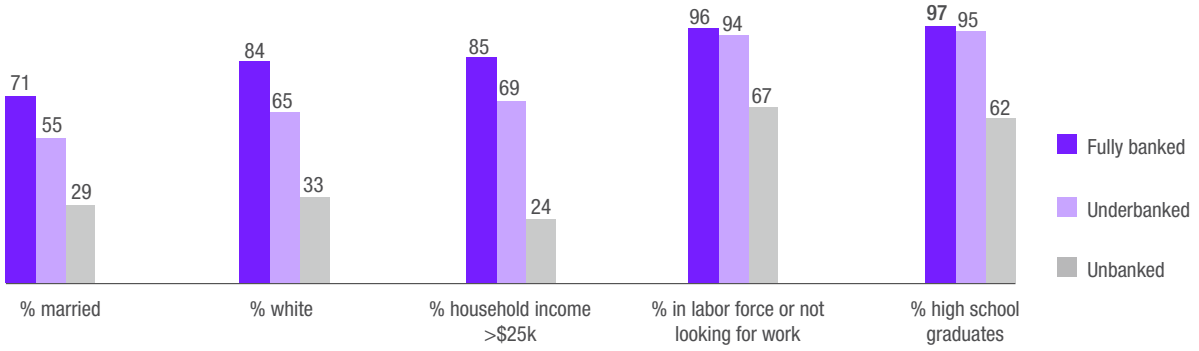
The three groups – fully banked, underbanked, and unbanked – have different demographic characteristics (Table 3). The two types of banked consumers (fully banked and underbanked) have somewhat similar characteristics (Figure 1). The underbanked are not very different from the fully banked, especially when compared with the unbanked, who are markedly different from the two banked groups.

Fully banked consumers tended to be older (average age 49.6) and more likely to be married (71%). More than 80% were white and fewer than one in seven had household income of less than U.S.$25,000 (Figure 1).

Compared with the fully banked, underbanked consumers were younger (average age 45.3) and just over half were married. Two-thirds were white and about one in three had household income less than U.S.$25,000 (Table 3). We estimated the effect of each demographic characteristic on banking status, while holding all other characteristics constant. Compared with fully banked consumers, African-Americans and Asian-Americans were more likely to be underbanked, as were consumers with income less than U.S.$25,000 and high school graduates (Appendix A, Table A.1). High-income consumers (income greater than U.S.$100,000) and homeowners were less likely to be underbanked.

Unbanked consumers differ substantially from the two banked groups. Compared with fully banked consumers, unbanked consumers were still younger

Figure 1: Banking status of U.S. consumers, by selected characteristics (percentage of consumers)



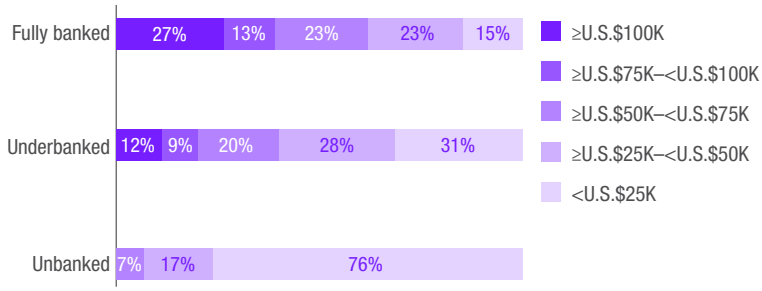
Source: 2014 SCPC, Federal Reserve Bank of Boston.
Note: For the unbanked, all demographic differences from the fully banked depicted here are statistically significant at the 5% level. For the underbanked, differences in marital status, race, and household income are also are statistically significant at the 5% level compared with the fully banked.

(average age 36.5), and fewer than one-third were married. Two-thirds were nonwhite and three in four had household income below U.S.$25,000. Unbanked consumers were far less likely to have graduated high school (62% are high school graduates) and to be unemployed and looking for work (67% were in the labor force or looking for work) (Table 3). In regression analysis, unemployed people, those with income below U.S.$25,000 or between U.S.$25,000 and U.S.$50,000, and African-Americans were more likely to be unbanked (Appendix A, Table A.1). Homeowners were less likely to be unbanked.

Income and banking status are related, a finding that is corroborated by regression results. Consumers with low income are more likely to be unbanked or underbanked. As Figure 2 shows, more than three-quarters of the unbanked had income below U.S.$25,000, compared with 31% of those who were underbanked, and 15% of those who are fully banked. In regression analysis holding other factors equal, income below U.S.$50,000 was significantly correlated with both unbanked and underbanked status, and income below U.S.$25,000 was strongly correlated with unbanked status see Appendix A for detailed regression results of all the demographic characteristics studied).

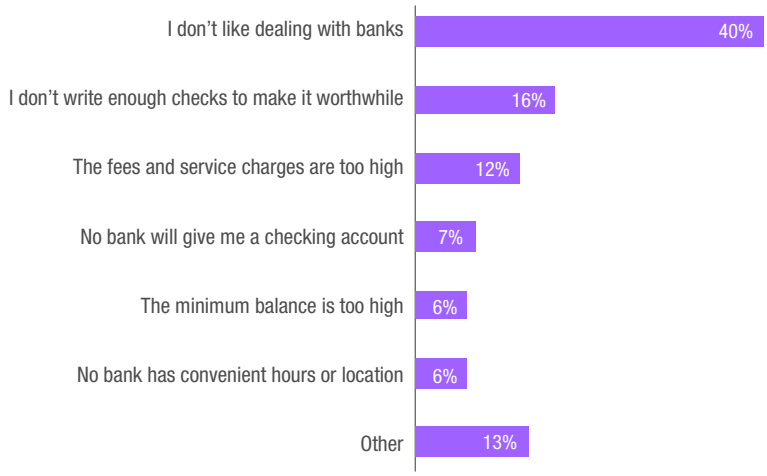
Income constraints are a factor in reasons consumers cite for being unbanked. Asked the primary reason they do not have a checking account, one-third of unbanked consumers cited reasons related to cost: that they did not write enough checks to make it worthwhile, that fees and service charges were too high, or that

Figure 2: U.S. consumers' income by banking status



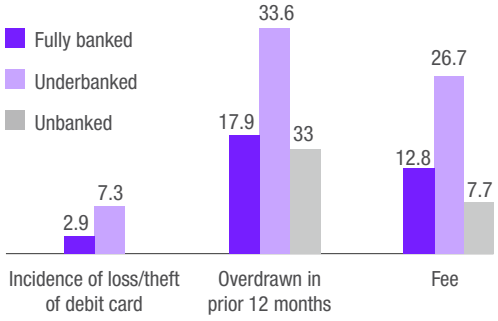
Source: 2014 SCPC, Federal Reserve Bank of Boston

Figure 3: Reasons given for not having a checking account



Source: 2014 SCPC, Federal Reserve Bank of Boston.
Note: Data in Figure 3 are unweighted

Figure 4: Percentage experiencing adverse events, by banking status

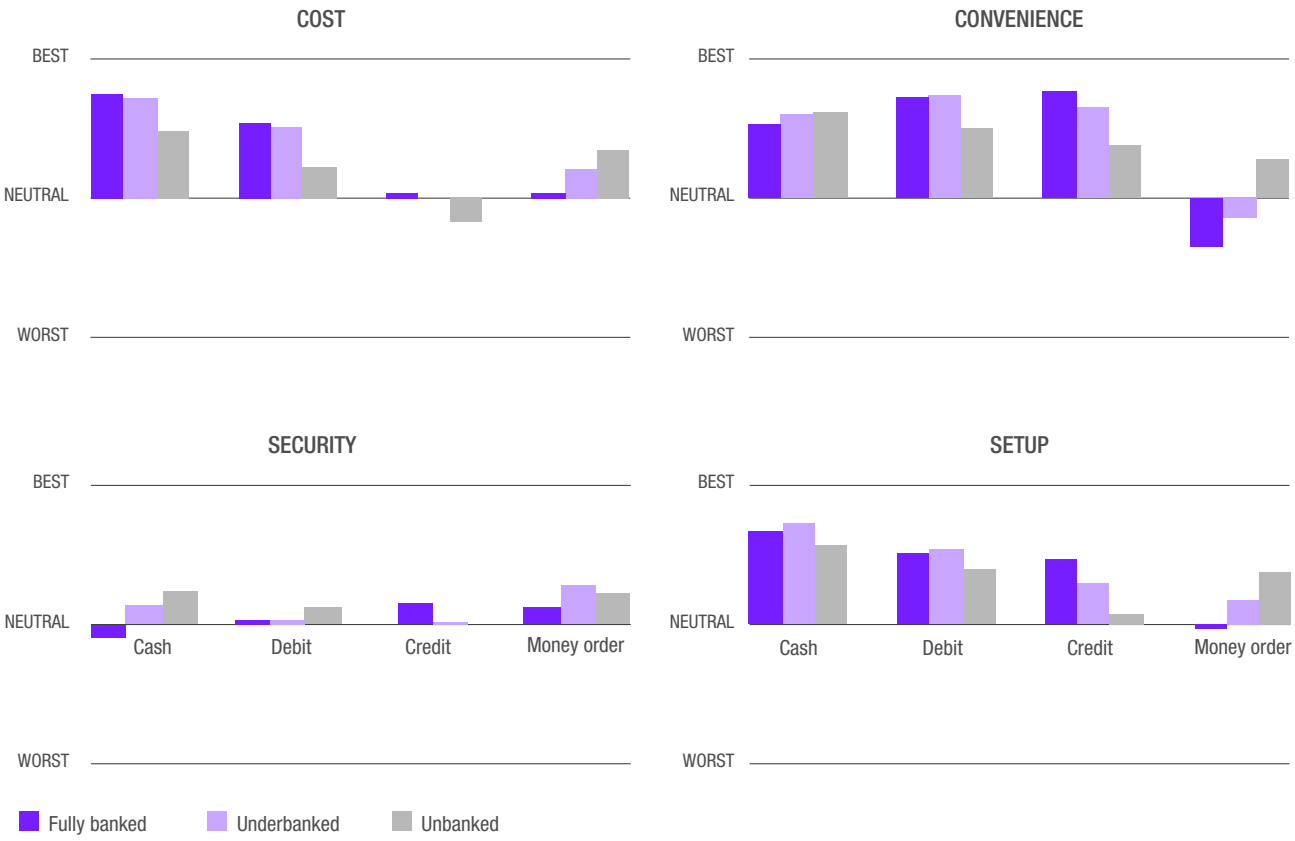


Source: 2014 SCPC, Federal Reserve Bank of Boston.

minimum balances were too high (Figure 3). These objections make sense, given that consumers with lower income would be more likely to face higher cost when obtaining banking services (for example, due to lower balances held in their accounts). Two in five unbanked consumers answered more generally, saying that they "don't like dealing with banks." This answer could encompass a whole range of interactions, including some related to income constraints. A small percentage reported that no bank would permit them to open a checking account. Consumers chose the primary reason for being unbanked, so it is possible that supply-side restrictions apply to other consumers as well. That is, the percentage of consumers who have no choice but to be unbanked may be understated.

Income was also significantly related to underbanked status. As noted above, the concept of being "underbanked" is not clear-cut. Compared with fully banked consumers, underbanked consumers were significantly more likely to have had income below U.S.$50,000. Note that 31% of underbanked consumers had income below U.S.$25,000, compared with 15% of fully banked consumers (Figure 2). Underbanked consumers were also more likely to have overdrawn an account in the 12 months ended in October 2014 (an event related to income constraints) and also to have paid a fee for being overdrawn (Figure 4). In addition, underbanked consumers were more likely to have experienced loss, theft, or fraud related to a debit card (7.3% compared with 2.9%) than were banked consumers. In the regression model (Appendix A), loss or theft of a debit card and having overdrawn an account in the past 12 months were also significantly associated with having underbanked status.

Figure 5: Average ratings of payment instruments, by banking status

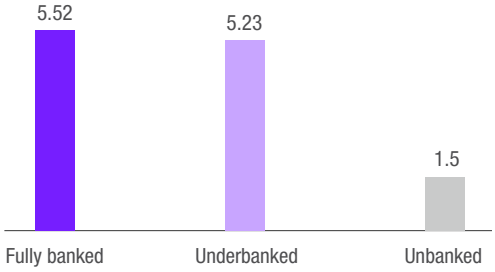


Source: 2014 SCPC, Federal Reserve Bank of Boston. Notes: in survey questionnaire "best" is the most positive assessment on a five-point scale. Cost: differences in cost ratings by unbanked versus fully banked consumers are significant for all instruments depicted here. Differences between cost ratings of money order by underbanked consumers and those by the fully banked are statistically significant. Convenience: differences in convenience ratings by unbanked and underbanked consumers compared with those of fully banked consumers are significant for debit, credit, and money order. Differences in convenience ratings for credit and money order are statistically significant by underbanked consumers when compared with those by fully banked consumers. Security: differences in security ratings by unbanked and underbanked consumers versus those by fully banked consumers are significant for cash. Underbanked consumers rate credit cards, prepaid cards, and money orders as significantly more secure than fully banked consumers do. Setup: differences in ratings by both underbanked and banked consumers for credit card (more difficult to set up) and money order (less difficult to set up) are significant when compared with those by fully banked consumers.

5. PAYMENT INSTRUMENT ASSESSMENTS

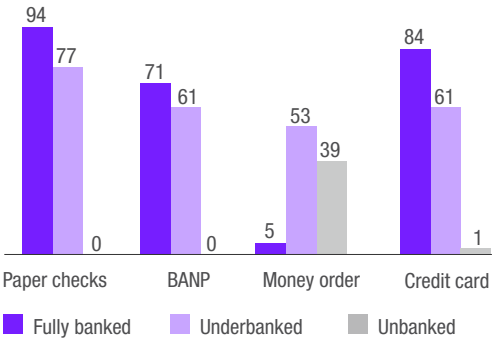
Unbanked consumers' nonspecific dislike of banks could flow through to assessments of payment instruments. Assessments of payment instrument characteristics have been found to affect payment behavior, with a follow-on effect on payment instrument adoption and use [Koulayev et al. (2016), Schuh and Stavins (2013, 2015b)]. We examined four characteristics that could affect payment instrument adoption or use: cost, convenience, security, and ease of setup. With "5" being the most positive assessment on a 1-to-5 scale (5 is shown as "best" and 1 is shown as "worst"), Figure 5 shows ratings for four instruments (paper checks and prepaid cards are omitted from the figure for clarity). Compared with fully banked consumers, underbanked and unbanked consumers offered generally less positive assessments of four mainstream payments instruments for cost and setup: cash, debit card, credit card, and (not shown) paper check. They see cash as being more convenient and more secure than do fully banked consumers. Compared with fully banked consumers, underbanked and unbanked consumers offered generally more positive assessments of money orders on all four characteristics. Prepaid ratings on all characteristics are generally similar for the three groups.

Figure 6: Average number of payment instrument types adopted by consumers (of eight available), by banking status



Source: 2014 SCPC, Federal Reserve Bank of Boston.
Note: money orders are excluded from this calculation. See text for explanation.

Figure 7: Percentage of consumers adopting payment instrument by banking status



Source: 2014 SCPC, Federal Reserve Bank of Boston. Note: Differences from adoption rates by the fully banked are statistically significant at the 5% level. For other instruments not shown (cash, debit card, prepaid card, and OBBP), there is no statistically significant difference in adoption rates of underbanked consumers compared with adoption rates of fully banked consumers. For prepaid cards, there is no statistically significant difference in adoption rates of unbanked consumers as compared to fully banked consumers. 100% of consumers in all three categories have adopted cash.

Unbanked consumers consistently rate the cost of mainstream payment instruments – including cash – more poorly than fully banked and underbanked consumers do (Figure 5). For cash, paper checks, debit cards, and credit cards, the differences in ratings by unbanked consumers compared to fully banked consumers for all four payment instruments are statistically significant. Both unbanked and underbanked consumers rate money orders as less costly than fully banked consumers do; these differences also are statistically significant.

Being less banked is correlated with seeing cash and money order as more convenient, and credit and debit as less convenient. Both underbanked and unbanked consumers rate credit as significantly less convenient and money order as significantly more convenient, compared with ratings by fully banked consumers. Unbanked consumers also rate debit cards significantly less convenient than fully banked consumers do. The three groups did not rate the convenience of checks, cash, or prepaid cards differently.

Unbanked and underbanked consumers both rated credit cards more poorly than banked consumers did for setup, defined as "the task of getting or setting up each payment method before you can use it" on a five-point scale from "very hard" to "very easy." Both groups rated money orders more favorably for setup, compared with fully banked consumers. Presumably, lack of familiarity with obtaining a credit card and familiarity with using money orders were factors in these ratings. The differences in these ratings are statistically significant.

Consumers frequently report that security is a very important, or the most important, attribute in evaluating a payment instrument. In each annual SCPC between 2008 and 2012, consumers ranked security as the most important characteristic of payments.[12] Several studies found security and identity theft important for payments adoption and use [Stavins (2013) and Kahn and Liñares-Zegarra (2015)]. Both underbanked and unbanked consumers viewed cash as significantly more secure, than fully banked consumers did.

Underbanked consumers also said prepaid cards and money orders were significantly more secure, compared with banked consumers security is the only characteristic for which a difference in assessment of prepaid cards was significant). Fully banked consumers rated both cash and prepaid cards negatively for security.

[12] Most respondents considered convenience to be most important in 2013. This question was omitted from the 2014 SCPC.

6. PAYMENT INSTRUMENT ADOPTION

6.1 Number of payment instruments adopted

By definition, unbanked consumers have a restricted choice of payment instruments. Their options are very limited. They do not have access to the four payment instruments linked to a bank account (paper checks, debit card, BANP, and OBBP). It is, therefore, not surprising that, of eight payment instruments (excluding money orders because money orders are part of the definition of being underbanked), the average unbanked consumer held just 1.5 payment instruments (Figure 6). Underbanked consumers may use quite a few payment instruments as they put together a mosaic of bank-linked products and nonbank products (for example, money order purchased from the U.S. Postal Service); there was only a small difference in the number of payment instruments adopted by these underbanked and fully banked consumers (5.52 for fully banked compared with 5.23 for underbanked). When money orders are included, the numbers of instruments adopted are 5.57 for fully banked, 5.77 for underbanked, and 1.87 for unbanked.

6.2 Adoption rates of individual payment instruments

As noted above, unbanked consumers have limited choice in the adoption of payment instruments. In addition, compared with fully banked consumers, unbanked consumers are more likely to have adopted money orders (39% compared with 5%) and less likely

Table 4: Credit card adoption, by banking status

% ADOPTING	FULLY BANKED	UNDERBANKED	UNBANKED
Credit or charge	84.4	60.6*	1.3*
Credit	84.2	60.2*	1.3*
Charge	6.8	4.8	1.3*
Median # of credit and/or charge cards	3 (3.58)	2 (2.37*)	0 (.16*)
Of adopters, percent revolving	55	66	NA

Source: 2014 SCPC.
Note: * indicates significantly different from the "fully banked" group at the 5% level.

Table 5: Demographic comparison, by banking status (percentage unless otherwise indicated)

		FULLY BANKED			UNDERBANKED			UNBANKED		
Percentage holding one of three most popular portfolios, by banking status		42			66			80		
Percentage holding portfolio, by banking status		19.3	15.3	7.7	24.5	23.7	17.3	29	27.8	22.7
Number of payment instruments held		7	6	6	6	7	8	1	2	3
Payment instruments linked to bank account	Check	✓	✓	✓	✓	✓	✓			
	Debit card	✓	✓	✓	✓	✓	✓			
	BANP	✓	✓	✓	✓	✓	✓			
	OBBP	✓	✓	✓	✓	✓	✓			
Cash		✓	✓	✓	✓	✓	✓	✓	✓	✓
Prepaid card		✓		✓		✓	✓		✓	✓
Money order							✓			✓
Credit card		✓	✓	✓	✓	✓	✓			

Source: 2014 SCPC, Federal Reserve Bank of Boston. 5% level. Results are weighted.

Table 6: Cash management, by banking status

	FULLY BANKED	UNDERBANKED	UNBANKED
Cash holdings (U.S.$)	230.89	166.18*	117.68*
Cash on person	60.11	58.41	40.63
Cash stored elsewhere	176.20	112.08*	78.96*
Cash holdings (as percentage of weekly income)	25.8	26.3	70.4
Number of withdrawals per month	5.1	7.9*	4.8
% check cashing store is primary	0.5	2.1	7.3*

Source: 2014 Survey of Consumer Payment Choice. Note: * indicates significantly different from the "fully banked" group at the 5% level. Percentage of weekly income computed using the midpoint of the annual income ranges described in footnote 18.

to have adopted credit cards (1% compared with 84%). Adoption rates of prepaid cards (all types, including general purpose reloadable (GPR) prepaid cards and gift cards among other types) are about the same as those of banked consumers.

Compared with fully banked consumers, underbanked consumers are less likely to have adopted two instruments linked to a checking account: paper checks[13] (77% compared with 94% for fully banked) and BANP (61% compared with 71% percent for fully banked, Figure 7). In addition, as expected, they are more likely to have adopted money orders, since purchasing a money order from a nonbank is among the criteria for being classified as underbanked.[14] Of note, underbanked consumers have less access to credit for day-to-day spending than the fully banked do; 61% have one or more credit cards compared with 84% for the fully banked (Table 4).[15]

6.3 Portfolios of payment instruments adopted

The mix of payment instruments adopted by consumers varies quite a bit; for the 1,809 Rand ALP respondents to the 2014 SCPC, there were 117 unique portfolios of payment instruments.[16] Fully banked consumers exhibited the most variety in their choices, followed by the underbanked and then the unbanked. The shares of consumers adopting each of the three most popular portfolios by banking status reflect, in part, the fewer choices available to unbanked consumers (Table 5). It is important to note, however, that more consumers fall into the fully banked category; this larger number of consumers could be another factor affecting the large number of portfolio mixes chosen by the fully banked.

7. ALTERNATIVES TO BANK ACCOUNTS FOR HOLDING ASSETS

As alternatives to a bank account, consumers may choose to hold funds as cash, in nonbank payments accounts, or on prepaid cards.[17]

Nearly all consumers have adopted cash, defined as using cash at least once in the prior 12 months or having some cash on person or property. While underbanked and unbanked consumers have significantly less cash on hand than fully banked consumers, this is likely related to their lower income, as discussed above. Taking income into account, unbanked consumers hold 70% of their weekly income in cash, compared with about 26% for fully banked and underbanked consumers (Table 6).[18]

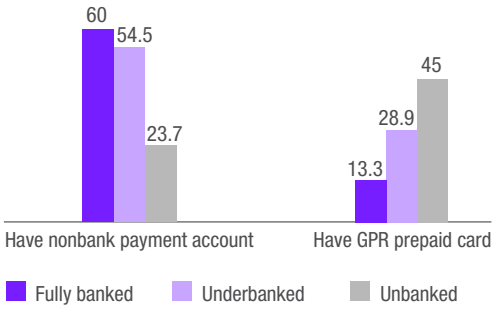
[13] Defined as currently having blank, unused checks or having written a paper check in the 12 months ending in October 2014.
[14] Consumers also may purchase money orders from banks.
[15] The statistical hypotheses of no difference in the adoption rates of checks, BANP, and credit cards between fully banked and underbanked consumers can each be rejected at the 95% significance level.
[16] A "unique portfolio" is a particular combination of payment instruments. For example, one unique portfolio is "cash." Another is "check, debit card, BANP, OBBP, cash, prepaid card, money order, credit card."
[17] Money orders are omitted from this discussion because owning a money order from a nonbank is one activity that satisfies the criteria for being underbanked.
[18] Percentage of weekly income computed using the midpoint of the following annual income ranges: <U.S.$5,000, U.S.$5,000–U.S.$7,499, U.S.$7,500–U.S.$9,999, U.S.$10,000–U.S.$12,499, U.S.$12,500–U.S.$14,999, U.S.$15,000–U.S.$19,999, U.S.$20,000–U.S.$24,999, U.S.$25,000–U.S.$29,999, U.S.$30,000–U.S.$34,999, U.S.$35,000–U.S.$39,999, U.S.$40,000–U.S.$49,999, U.S.$50,000–U.S.$59,999, U.S.$60,000–U.S.$74,999, U.S.$75,000–U.S.$99,999, U.S.$100,000–U.S.$124,999, U.S.$125,000–U.S.$199,999, ≥U.S.$200,000. Data are weighted.

Table 7: GPR prepaid card adoption, by banking status

ADOPTION RATES	FULLY BANKED	UNDERBANKED	UNBANKED
Included in the definition of GPR prepaid cards			
Other general-purpose prepaid card (cards not reported in specific categories below)	11.6	18.9*	20.4
Direct Express	0.00	2.5*	10.3*
EBT, WIC, SNAP, or TANF	6.0	13.7*	29.9
Other federal, state, or local government benefit card	0.1	4.5*	8.0*
Payroll card (for wages or salary)	0.4	2.3	3.0
At least one of any GPR type	13.3	28.9	45.0
Not included in the definition of GPR prepaid cards			
Gift card from a store, merchant, or website (examples: Home Depot, Target, Starbucks, iTunes)	32.0	22.1*	5.1*

Source: 2014 SCPC. Note: * indicates significantly different from the "fully banked" group at the 5% level.

Figure 8: Percentage of consumers adopting nonbank accounts, by banking status



Source: 2014 SCPC, Federal Reserve Bank of Boston.

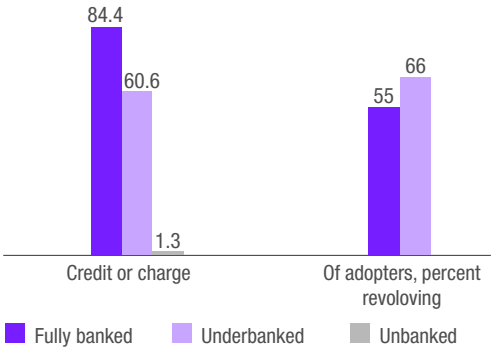
In nominal terms, underbanked and unbanked consumers withdraw more cash per month than banked consumers. Unbanked consumers withdraw U.S.$652, underbanked, U.S.$721, and fully banked, U.S.$486. Despite withdrawing more, these consumers have less cash on hand, as noted above, perhaps related to their heavy use of cash for payments (discussed below). For getting cash, unbanked consumers have fewer options than other consumers. For both fully banked and underbanked consumers, the most popular locations for getting cash are ATM machines and bank tellers. Unbanked consumers report a family member or friend and being paid in cash as their two most likely ways of getting cash. Unbanked consumers make greater use of check cashing stores than others: 7.3% report that check cashing stores are their primary source of cash, compared with 2.1% of underbanked consumers and 0.5% of banked consumers.[19]

Underbanked and unbanked consumers are significantly more likely to experience the loss or theft of cash, perhaps because they carry proportionately more cash or perhaps because they use it more often. Of fully banked consumers, 4.9% experienced the loss or theft of cash, compared with 12.6% of underbanked consumers and 14.4% of unbanked consumers.[20]

Consumers can also keep funds in nonbank accounts, such as PayPal, or store money on a prepaid card. Ownership of nonbank payment accounts (PayPal, etc.)[21] and GPR prepaid cards differs for the three groups (Figure 8). People who are unbanked are significantly less likely to have a nonbank payment account than are the fully banked or underbanked. Typically, these

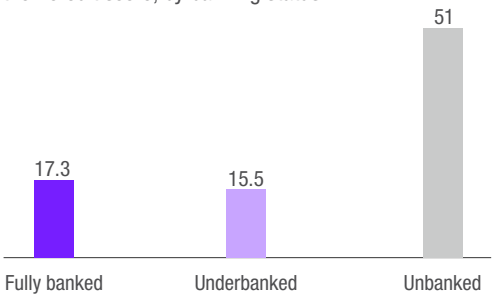
[19] Differences are statistically significant at the 5% level.
[20] Differences are statistically significant at the 5% level.
[21] The SCPC asks: "A nonbank online payment account is a payment service provided by a company that is not a bank. These services allow a consumer to send and receive money online, and pay for purchases or bills. Do you have an account at any of the following non-bank online payment services?"

Figure 9: Credit card adoption and revolving, by banking status



Source: 2014 SCPC, Federal Reserve Bank of Boston.

Figure 10: percentage of consumers who do not know their credit score, by banking status



Source: 2014 SCPC.

nonbank accounts are linked to traditional checking or savings accounts for depositing and withdrawing funds.

Compared with fully banked consumers, people who are unbanked are significantly more likely to have a GPR prepaid card (45% compared with 13% for fully banked consumers), as are people who are underbanked (29%) (Figure 8).[22] Adoption of GPR prepaid cards is defined as adoption of any of the following: (1) General-purpose prepaid card (has a logo from Visa, MasterCard, Discover, or American Express), (2) government benefit card including Direct Express; EBT, WIC, SNAP, or TANF; or other federal state, or local government benefit card, (3) payroll card (Table 6).

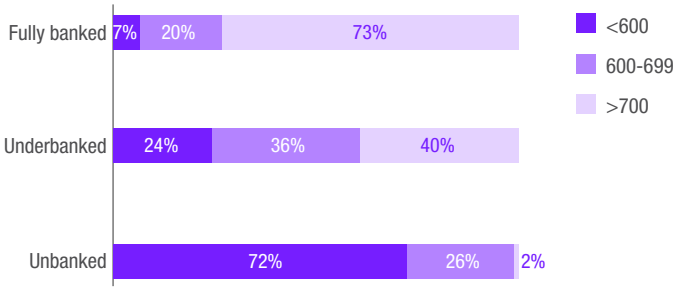
Looking in detail at individual types of cards, both unbanked and underbanked consumers are more likely to have prepaid cards for the receipt of government benefits and less likely to have gift cards, compared with fully banked consumers (Table 7).

8. ACCESS TO CREDIT

Both underbanked and unbanked consumers are less likely than fully banked consumers to have a credit or charge card. Almost no unbanked consumers have a credit card, just 1.3%. Majorities of fully banked and underbanked consumers have cards: 84.4% of the fully banked compared with 60.6% of the underbanked. Fully banked consumers own, on average, 3.6 credit cards – 50% more than underbanked consumers, who own 2.4. Among credit card adopters, the underbanked are significantly more likely than the fully banked to revolve on their cards: 66% of the underbanked credit card adopters revolve, compared with 55% of fully banked credit card adopters (Figure 9).

More than half of unbanked consumers do not know their credit scores, compared with 16 or 17% of underbanked and fully banked consumers (Figure 10). Of consumers who know their scores, three-quarters of unbanked consumers have poor scores (less than 600). About 1% of unbanked consumers report good or excellent scores (700 or more) compared with 40% of underbanked consumers and 73% of fully banked consumers (Figure 11). As noted above, fewer than 2% of unbanked consumers have a credit card, so it would be almost impossible for an unbanked consumer to develop a credit history that would lead to a high credit score.

Figure 11: Self-reported credit scores, percentage of consumers by banking status



Source: 2014 SCPC. Note: Consumers who answered "I don't know" are omitted.

[22] For additional discussion of GPR prepaid card holders who do not have checking accounts, see Greene and Shy (2015).

Table 8: Shares of transaction types, by banking status (percentage)

TRANSACTION TYPE	FULLY BANKED	UNDERBANKED	UNBANKED
Retail in person	35.2	31.7*	41.9
Services in person	22.0	20.8	19.1
Bill pay in person/by mail or phone	11.3	13.2*	27.2*
Bill pay online	10.6	10.5	0.0*
Bill pay automatic	11.3	11.3	.2*
Retail online	5.7	5.0	2.6*
Person to person	3.9	7.5*	8.9*

Source: 2014 SCPC. Note: *Shares are significantly different from shares of transaction type by fully banked consumers.

9. PAYMENT INSTRUMENT USE

Compared with fully banked and underbanked consumers, unbanked consumers make few payments per month, 28 payments versus about 70 for the other two groups.

Between 30% and 40% of the payments of all U.S. consumers are for retail goods and around 20% of their payments are for retail services. For bill payments, behavior diverges, with fully and underbanked consumers making two-thirds of their bill payments automatically or online, while unbanked consumers make essentially all their bill payments by mail, phone, or in person. Fees paid for alternative financial services are often cited as one cost of being unbanked; another is the time required to pay bills or arrange for financial services in person. Unbanked consumers also make a smaller share of retail online purchases than others do (Table 8).

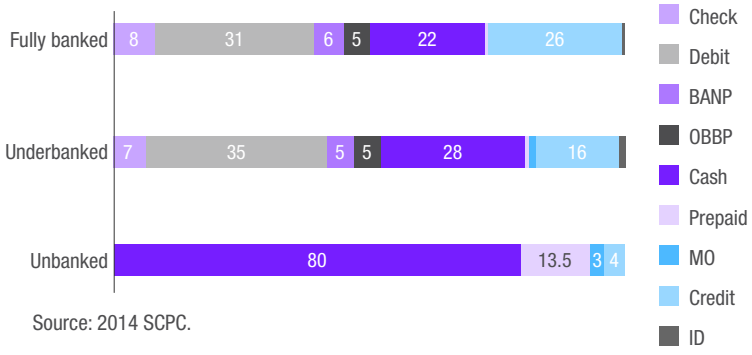
The relatively large shares of person-to-person payments made by both unbanked and underbanked consumers reflect their position outside the financial mainstream. This finding is similar to qualitative research by the CSFI, which has found that "casual lending and borrowing money from friends and family is common" [Tescher et al. (2007)]. In the 2014 SCPC, 8.9% of payments by unbanked consumers were made to another person, compared with 7.5% of underbanked consumers and 3.9% of fully banked consumers' payments.

Unbanked consumers use cash for four out of five of their payments and prepaid cards for most other payments (Figure 12). As noted above, they are unlikely to have a credit card in addition to lacking the four payment instruments linked to a bank account.

In contrast, cash payments of fully banked consumers represent a much smaller share of their payments (22%), prepaid cards an even smaller share (0.5%), and money orders also a share equal to less than 1% of their payments. The fully banked use credit cards for more than one-quarter of their payments.

Underbanked consumers rely more on cash than fully banked consumers do; the underbanked use cash for almost 30% of payments. They also rely more on debit cards and money orders. Compared with the fully banked, they use credit less and are less likely to write a check. Like the fully banked, they rarely use prepaid cards.

Figure 12: Self-reported credit scores, percentage of consumers by banking status



Source: 2014 SCPC.

10. SUMMARY

Looking at U.S. consumers by banking status (fully banked, underbanked, and unbanked), we find differences in income distribution, demographic characteristics, and payment behavior. Lower income is correlated with being un- or underbanked, with consumers with the lowest income most likely to be unbanked. The strong association with income indicates that consumers' stated preferences and reasons for being underbanked may be constrained by their income levels. Race and education also are associated with banking status.

Underbanked and fully banked consumers are fairly similar in their payment behavior. Each group makes about half of all payments (by number) using payment instruments linked to a bank account. Unbanked consumers rely heavily on cash; 80% of their payments are in cash. Reliance on cash means that unbanked consumers pay almost all bills in person or by mail or phone; consumers with a bank account (fully and underbanked) pay two-thirds of bills online or automatically.

Unbanked status is explicitly defined; being underbanked is a fuzzier concept. Further research and survey modifications would be needed to understand underbanked consumers' motivations and constraints more clearly as well as to define their status more precisely.

APPENDIX A

We examine the effects of demographics and income on underbanked or unbanked status (as opposed to "fully banked") using probit regressions. The first column reports the results of a probit regression including observations for the fully banked and the unbanked, with the dependent variable being a 0/1 indicator for being unbanked. This regression excludes the underbanked for the sake of obtaining a strict comparison between the unbanked and the fully banked. Control variables include respondent demographics, adverse experience, and financial responsibility within the household. For the unbanked regression, the top three income categories (U.S.$50,000–U.S.$74,999, U.S.$75,000–U.S.$99,999, >U.S.$100,000) are collapsed into one, due to lack of observations. Responses associated with experience with bankruptcy, debit card theft, and credit card account closure were also excluded from the unbanked regression due to lack of observations.

The second column reports the results of a probit regression including observations for the fully banked and the underbanked, with the dependent variable a 0/1 indicator for being underbanked. This regression excludes the unbanked for the sake of obtaining a strict comparison between the fully banked and the underbanked. Control variables include respondent demographics, adverse experience, and financial responsibility within the household.

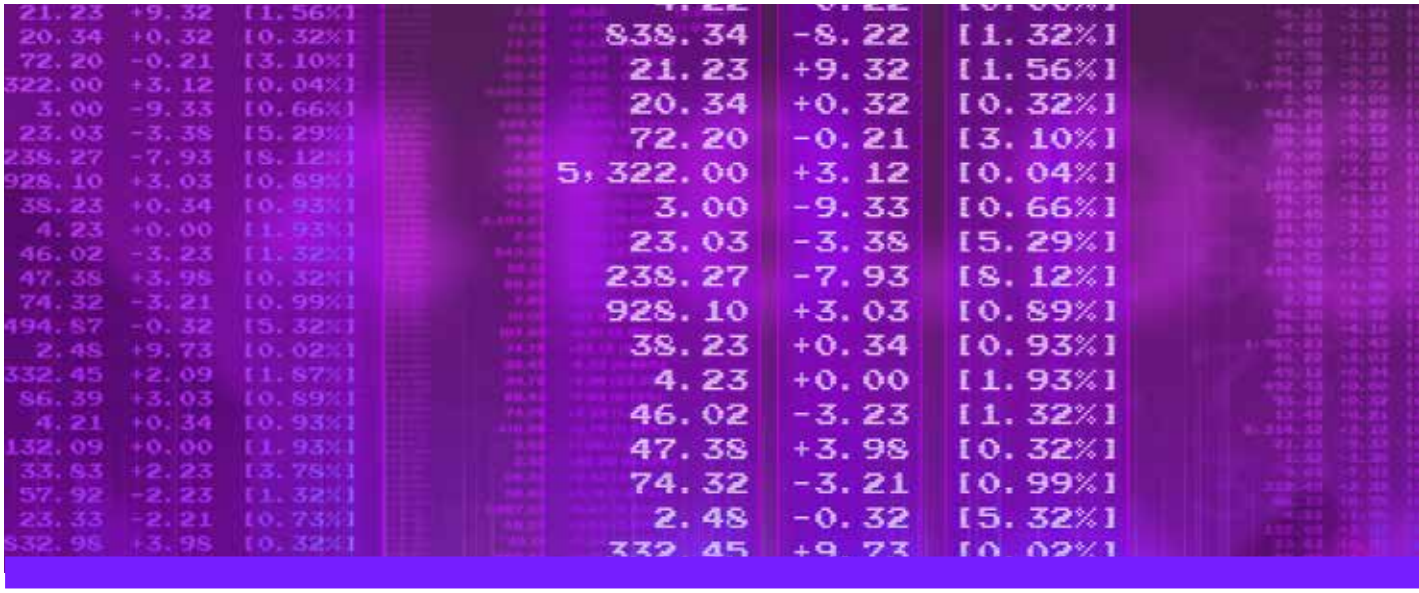
Reference groups for each demographic category are as follows: age 35–44, male, white, non-Latino, college graduate, never married, born in the United States, income U.S.$50,000–U.S.$74,999 (underbanked), income >U.S.$50,000 (unbanked), employed, resident of the Northeast, equally shared bill pay responsibilities.

Table A.1: Probit regressions, effects of demographics and income on underbanked or underbanked status

INDEPENDENT VARIABLES		UNDERBANKED	UNBANKED
Age	<25	0.01	-0.56
	25 – 34	-0.08	-0.27
	45 – 54	-0.14	-0.13
	55 – 64	-0.10	-0.42
	≥ 65	-0.23	-0.63
Gender	Female	-0.05	-0.42[c]
Race	Black	0.72[a]	1.03[a]
	Asian	0.71[a]	0.25
	Other	0.23	0.20
Ethnicity	Latino	0.15	0.46
Education	Less than high school	0.40	2.03[a]
	High school	0.34[a]	1.39[a]
	Some College	0.04	1.02[b]
	Postgraduate	0.14	1.15[b]
Marital Status	Married	-0.19	-0.10
	Divorced	-0.09	-0.09
	Separated	-0.30	-0.30
	Widowed	-0.68	-0.68
Nationality	Immigrant	-0.27	-0.31
Income	<U.S.$25,000	0.31[b]	1.37[a]
	U.S.$25,000 – U.S.$49,999	0.65[b]	0.65[b]
	U.S.$75,000 – U.S.$99,999	-0.13	NA
	≥U.S.$100,000	-0.30[b]	NA
Employment Status	Retired	-0.07	-0.37
	Disabled	0.18	0.16
	Unemployed	-0.17	0.59[c]
	Homemaker	-0.27	-0.03
	Other	0.16	0.75

INDEPENDENT VARIABLES		UNDERBANKED	UNBANKED
Geographic Region	Mid-Atlantic	0.14	-0.60
	East North Central	-0.16	-0.16
	West North Central	0.18	0.18
	South Atlantic	0.26	0.26
	East South Central	0.00	0.00
	West South Central	-0.68	-0.68
	Mountain	0.19	0.19
	Pacific	0.05	0.05
Bill pay financial responsibility	None or almost none	-0.09	0.30
	Some	0.08	-0.61
	Most	0.32[c]	-0.11
	All or almost all	0.10	-0.19
Household size	Household size	0.04	0.05
Home ownership	Owns Home	-0.27[a]	-1.00[a]
Financial adversity	Bankruptcy within the last year	0.36	NA
	Bankruptcy within the last 7 years	0.32[c]	-0.32
	Foreclosure within the last year	-0.66	0.70
	Foreclosure within the last 7 years	-0.14	-0.53
	Job loss within the last year	0.17	-0.54
	Overdraft within the last year	0.30[a]	-0.33
	Stolen debit card in the last year	0.36[b]	NA
	Credit card account closed in the last year	-0.00	NA
	N	1663	1332
	Pseudo R-squared	0.14	0.52

Source: 2014 SCPC. Note: a indicates significance at the 1% level, b indicates significance at the 5% level, and c indicates significance at the 1% level. Note: The variables representing income of U.S.$75,000–U.S.$99,999, income greater than U.S.$100,000, and bankruptcy within the last year were excluded from the unbanked regression due to lack of observations.



References

Angrisani, M., K. Foster, and M. Hitczenko, 2016, "The 2014 survey of consumer payment choice: technical appendix," Federal Reserve Bank of Boston Research Data Report 16-04

Burhouse, S., K. Chu, R. Goodstein, J. Northwood, Y. Osaki, D. Sharma, 2014, "2013 FDIC national survey of unbanked and underbanked households," Federal Deposit Insurance Corporation, http://bit.ly/1ALvL0n

Burhouse, S., K. Chu, K. Ernst, R. Goodstein, A. Lloro, G. Lyons, J. Northwood, Y. Osaki, S. Rhine, D. Sharma, and J. Weinstein, 2016, "2015 FDIC national survey of unbanked and underbanked households," Federal Deposit Insurance Corporation, http://bit.ly/2vPFur9

Demirgüç-Kunt, A., L. Klapper, D. Singer, and P. Van Oudheusden, 2015, "The global Findex database 2014: measuring financial inclusion around the world," World Bank Policy Research Working Paper 7255

FDIC, 2014, "Assessing the economic inclusion potential of mobile financial services," Federal Deposit Insurance Corporation, June 30

Greene, C., S. Schuh, and J. Stavins, 2016, "The 2014 survey of consumer payment choice: summary results," Federal Reserve Bank of Boston Research Data Report 16-3

Greene, C., and O. Shy, 2015, "How are U.S. consumers using general purpose reloadable prepaid cards? Are they being used as substitutes for checking accounts?" Federal Reserve Bank of Boston Research Data Report 15-3

Gutman, A., T. Garon, J. Hogarth, and R. Schneider, 2015, "Understanding and improving consumer financial health in America," Center for Financial Services Innovation

Hogarth, J. M., C. E. Anguelov, and J. Lee, 2005, "Who has a bank account? Exploring changes over time, 1989–2001," Journal of Family and Economic 26:1, 7-30

Kahn, C. M., and J. M. Liñares-Zegarra, 2015, "Identity theft and consumer payment choice: does security really matter?" Journal of Financial Services Research 50:1, 1–39

Koulayev, S., M. Rysman, S. Schuh, and J. Stavins, 2016, "Explaining adoption and use of payment instruments by U.S. consumers," RAND Journal of Economics 47:2, 293–325

Schuh, S., and J. Stavins, 2010, "Why are (some) consumers (still) writing paper checks?" Journal of Banking and Finance 34:8, 1745–1758

Schuh, S., and J. Stavins, 2013, "How consumers pay: adoption and use of payments," Accounting and Finance Research 2:2, 1-21

Schuh, S., and J. Stavins, 2015a, "The 2013 survey of consumer Payment Choice," Federal Reserve Bank of Boston Research Data Report 15-4

Schuh, S., and J. Stavins, 2015b, "How do speed and security influence consumers' payment behavior?" Contemporary Economic Policy 34:4, 595–613

Stavins, J., 2013, "Security of retail payments: the new strategic objective," Federal Reserve Bank of Boston Public Policy Discussion Paper 13-9

Tescher, J., E. Sawady, and S. Kutner, 2007, "The power of experience in understanding the underbanked market," Center for Financial Services Innovation

White House Council of Economic Advisers, 2016, "Financial inclusion in the United States," June

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, LinkedIn and Xing.**

WORLDWIDE OFFICES

Bangalore
Bratislava
Brussels
Chicago
Dallas
Dusseldorf
Edinburgh
Frankfurt
Geneva
Hong Kong
Houston
Kuala Lumpur
London
New York
Orlando
Paris
Pune
São Paulo
Singapore
Stockholm
Toronto
Vienna
Warsaw
Washington, DC
Zurich

CAPCO.COM

CAPCO