Social Impact Bonds (SIBs): Analyzing Efficiency and Legitimacy Motivations Using Systematic Review Methods and Social Media Analysis



Vanessa Picker

Department of Social Policy and Intervention

Jesus College

University of Oxford

Michaelmas Term 2022

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Social Intervention and Policy Evaluation in the Social Sciences Division at the University of Oxford

99, 023 words (excluding bibliography)

Abstract

Policymakers frequently endorse evidence-informed policy design. But, in the context of public management reform, the most technically suitable innovations are often not the ones that become widely and deeply adopted. Legitimate (i.e. socially 'acceptable') reforms can spread greatly within and across public sectors, despite being inefficient; and efficient reforms may nonetheless remain unimplemented if they lack the legitimacy necessary to gain support. Inefficient reforms may even become "institutionalized", which means they are regarded as having social value beyond their technical value. At best, the result is ineffectual public management reforms, with attendant opportunity costs; at worst, poorly-targeted interventions directly harm public service delivery and policy outcomes.

In between these widespread and *un*-spread management initiatives, however, are those reforms that gain some traction among governments, yet never really "take-off" to become fully institutionalized. Social Impact Bonds (SIBs) are a recent example of this. A SIB is a type of payment by results (PbR) contracting, in which the commissioner pays for outcomes achieved rather than services delivered, using upfront capital from private investors who earn a financial return if the intervention is successful. As of 2023, more than twelve years after the first SIB was launched in Britain, there have been 276 SIBs in 38 countries, with approximately $745 million USD raised in upfront capital and 1.7 million beneficiaries reached (GO Lab, 2023a). This growth is far less than initially expected; in recent years (from 2019 onwards), the usage of SIBs has decreased markedly. In 2021, only 19 SIBs were launched (raising approximately $25 million USD), compared to 44 SIBs in 2018 (with $54 million USD raised). Thus, the spread of SIBs has been "wide" but "thin" – i.e., to many places and sectors, but only to a very limited number of organizations or programs in each.

What explains this truncated reform diffusion? Why did SIBS rapidly appear in successive jurisdictions, but penetrate those public sectors barely at all? And what does this "in-betweener" category of innovation, neither widespread nor *un*-spread, tell us about the combined effect of instrumental and institutional, or efficiency and legitimacy, pressures for public management reform, and the role of evidence in decision-making? The empirical chapters of this thesis suggest that this reform has been truncated and only 'thinly' diffused because it has often been adopted for institutional reasons. The legitimacy benefits have caused decision makers to interrogate the reform idea with lower thresholds of rigour than they otherwise would. Yet, legitimacy benefits can only propel the innovation forward so far and for so long without concrete evidence of effectiveness. Thus SIBs usage has declined, in recent years, with Development Impact Bonds (DIBs) being introduced (and gaining popularity), likely in a deliberate attempt to re-package the idea as novel.

Building on seminal works in organizational sociology, this thesis shows how efficiency and legitimacy exercise varying influence over reform processes as levels of uncertainty change over time regarding the effectiveness of an innovation. I use an effectiveness systematic review of SIBs to gauge the changing state of academic knowledge about the benefits of SIBs during 2010-22, as well as a qualitative systematic review to identify broader drivers for the adoption of the reform. I perform sentiment analysis and network analysis on social media data to understand changing practitioner understanding of the reform between 2010-2020 and use framework analysis to understand how rhetorical and discursive techniques have been applied, in this same period, to frame the SIB idea as a worthwhile reform to pursue. Overall, these analyses demonstrate that legitimation is central to the early-stage diffusion of management innovations that, ultimately, spread widely but only thinly.

Acknowledgments

I arrived in Oxford thinking that I would spend one short, intense year here (for the MSc in Evidence-Based Social Intervention & Policy Evaluation) and return to my life in Australia. But here I am, many years on, writing the acknowledgments section for the DPhil thesis I never expected to write, let alone, submit. It was my long-term mentor, Allyson Essex, who gave me the confidence to apply for and accept a position at Oxford. It was Allyson who re-assured me that while pursuing a DPhil would be a significant change of plans, it would absolutely be worth it. And it was Allyson who observed that I am no longer the person I was when I left Australia to embark on my Oxford journey; I cannot and should not try to 'box' myself back in and I should accept that I have changed in many ways, hopefully for the better.

In my time at Oxford, I have had some of the best and worst times of my life. I have seen the best and worst of society, for instance, with cutting edge research and ideas emerging, yet backward and often exclusionary traditions and policies remaining. As somebody who has actively campaigned for gender equality at Oxford and beyond, I am well aware that the fight for women to be awarded degrees has been a very long one. For decades, women could attend lectures and take exams but not receive a degree or graduate. Even when women were finally permitted to matriculate and graduate, resistance remained, for instance, with quotas and restrictions put in place to limit the number of women. Against this backdrop, women became increasingly politically active, fighting for the rights that we often take for granted today. I am thus incredibly inspired by and grateful to all the women who have come before me, who fought long and hard for equal rights. I realize that despite progress, there is still a very long way to go, especially in Oxford. I hope that in my time

here, I have made a small difference in the ongoing fight for gender equality, while also helping to address other forms of injustice and inequity.

To the "EBSIPE fam" (i.e. my amazing MSc cohort); I could not be any more grateful for all of the memorable times we had together, which will always stand out to me, when I reflect back on my time at Oxford. While we are now spread out across the world, and the MSc days seem like a lifetime ago, I still feel very indebted to everybody and I feel as though I could not have shared my initial times at Oxford, with a better, more inspiring group of people. A challenging part of being here for the long-haul and completing a DPhil is that cohorts (and close friends) come and go; but the silver lining of this, has been that I've been fortunate enough to meet a new bunch of inspiring people every year. So, to the many cohorts of (13) Ship street residents, Jesus College MCR members and DSPI DPhil students, thank you also, for making my time here so fun and memorable.

When I first received an offer to Oxford, I was excited yet aware that accepting it would only be possible, with significant financial support. So, the endless hours of scholarship research and applications began. Without the extremely generous support of the Rotary Foundation, I would not have been able to undertake the EBSIPE MSc; and I certainly would not have been able to progress to an Oxford DPhil. But it is not just financial support that comes with a Rotary scholarship. I also owe my thanks to the many Rotarians that have welcomed me and made my time here so special. In particular, Martin and Annie Phillips, who were an invaluable support during my initial year here and beyond. Also, Richard Kirby (and Helen and Simon) who have become and will forever be my English family and who have been there for me in every way that they could be. To the Clarendon Fund, thank you for the financial support throughout my DPhil. As with the Rotary scholarship, the Clarendon

scholarship has always felt like much more than financial support. I'm also grateful for all of the invaluable events, social activities and friendships that have stemmed from this.

I could not write an acknowledgements section without explicitly mentioning my nan, Dot Picker, who sadly passed away shortly after my arrival in Oxford but has stayed on my mind throughout my time here. While she spent her life in outback Australia, far removed from the world of Oxford, my nan certainly appreciated how much of a privilege it was for me to have the opportunity to study here. If she was still here today, I suspect she would be bragging about my Oxford DPhil as much as she always boasted about having bowled me out in a game of backyard cricket (and I would still be disputing whether it was a fair delivery or a no-ball). I hope that I have made her proud and that I can use the learnings of my time here, to help others, as she always strived to do. To the rest of my family and friends back in Australia, thank you for supporting me from afar throughout my time at Oxford; and for always welcoming me back, on my visits, with such open arms. To my mum, in particular, thank you for the endless support you have provided.

I also owe an enormous thanks to my two supervisors, Dave Humphreys and Thomas Elston. From the very early days of my MSc, Dave encouraged me to think big and to consider applying to the DPhil program. Without his unwavering support and encouragement, I would not be where I am today. Dave has been an incredible support, across my many years here and in addition to pushing me academically, has been someone I could always turn to, in the good and bad times, for support. As my topic grew and broadened, it became clear that it would be beneficial to involve a second supervisor. I could not think of a more supportive and inspiring person, to have joined my supervisory team, than Thomas Elston. Thomas has been incredibly generous with his time, has always encouraged me in my many moments of

doubt and has helped me to think creatively and find an interesting “hook” in the sometimes-overwhelming body of public administration literature. I always tell people that my supervision ‘dynamic’ has worked so well because underneath all the formalities, it has often felt like we are good friends who enjoy catching up and that this has led to an amazing flow of ideas and conversation. I hope that our supervisory relationship can transition to a long-lasting friendship in which we can continue exchanging ideas and working together to cause positive change.

Table of Contents

Abstract ***2***

Acknowledgments ***4***

List of tables ***12***

List of figures ***14***

List of abbreviations ***17***

Chapter 1: Introduction ***19***

1.1 The emergence of SIBs **23**

1.2 A truncated diffusion trajectory **27**

1.3 Thesis outline **28**

Background 28

Empirical research 29

Interpretations and Contributions 31

1.4 Conclusion **32**

Chapter 2: Literature Review ***34***

2.1 Introduction **34**

2.2 Definition, types and stages of innovation **34**

2.3 Drivers of innovation **40**

2.3.1 Why organizations innovate 41

2.3.2 Why organizations do not innovate 50

2.3.3 Why organizations abandon innovations 51

2.4 Outcomes of innovation **52**

2.4.1 Social interventions often prove to be disappointing and are poorly measured (if at all) 52

2.4.2 Pursuing an innovation may be beneficial regardless of results 53

2.4.3 There are high opportunity costs associated with innovation 54

2.4.4 There is a strong selection bias towards successful ideas 55

2.4.5 Pro innovation bias emerged in the early literature 56

2.4.6 Researchers have now introduced an anti-innovation bias 57

2.5 Conclusion **57**

Chapter 3: Theoretical Framework ***59***

3.1 Overview **59**

3.2 Central Hypothesis **59**

3.3 The instrumental and institutional views **63**

3.3.1 Instrumental view 63

3.3.2 Institutional view 68

3.4 Applying the two alternate views to the context of SIBs **73**

Instrumental view 74

Institutional view 75

3.5 Conclusion **76**

Chapter 4: Research Methodology ***78***

4.1 Introduction **78**

4.2 Existing approaches and their shortcomings 78
4.2.1 Two-stage model of diffusion 78
4.2.2 Citation Tracking Approach 81

4.4 Key Methodological Components 85
4.4.1 Effectiveness systematic review 89
4.4.2 Qualitative systematic review 93
4.4.3 Quantitative and Qualitative Analysis of Social Media Data 97

4.4 Ethical considerations 102

4.5 Conclusion 103

Chapter 5: A Systematic Review of Effectiveness 106

5.1 Introduction 106

5.2 Overview of methods 107
Search strategy 109

5.3 Results 113
5.3.1 Search Results 113
5.3.2 Excluded studies 115
5.3.3 Included studies 116
5.3.4 Effectiveness evidence 123
5.3.5 Risk of bias in included studies & effects of interventions 138

5.4 Discussion 142

5.5 Conclusion 151

Chapter 6: Qualitative Systematic Review 153

6.1 Introduction 153

6.2 Overview of methods 155
Search strategy 156
Qualitative synthesis 161
Results 163

6.3 Findings 164
6.3.1 Characteristics of included studies 165
6.3.2 Quality appraisal 186
6.3.3 Excluded studies 186
6.3.4 Thematic Synthesis: Motivations for Pursuing SIBs 186
6.3.5 Contribution of individual studies to the identification of themes 200
6.3.6 Analytical themes 200

6.4 Discussion 206

6.5 Conclusion 211

Chapter 7: Sentiment Analysis and Network Analysis of Social Media Data 213

7.1 Introduction 213

7.2 Broad Trends Regarding the Development of the SIB Innovation 215

7.3 Hypotheses Development 225
R1: How have sentiments regarding SIBs changed over time on Twitter, and to what extent do the sentiments expressed in the tweets express emerging (inconclusive) research findings? 225
R2: Which stakeholders have played a central role in communicating and promoting the SIB idea in online forums such as Twitter? 226

7.4 Overview of Methods....227
Descriptive statistics:....229
Sentiment analysis....230
Social network analysis....231

7.5 Findings....232
R1: How have sentiments regarding SIBs changed over time on Twitter, and to what extent do the sentiments expressed in the tweets reflect those expressed in the academic literature?....232
R2: Which stakeholders have played a central role in communicating and promoting the SIB idea in online forums such as Twitter?....244

7.6 Discussion....253

7.7 Conclusion....258

Chapter 8: Framework Analysis of Social Media Data....260

8.1 Introduction....260

8.2 Hypotheses Development....262
R1: What persuasive techniques have been used by involved stakeholders to frame the SIB funding model as a 'legitimate' reform to pursue over time?....262
R2: What discourses can be identified in the highly shared tweets; and to what extent have particular discourses dominated/others been drowned out?....262

8.3 Overview of Methods....263
Overarching approach - framework analysis....264
Strategies of persuasion framework – random (10%) sample....265
Discourses framework – purposive sample....266
Additional analysis - measuring COI....268

8.4 Ethical considerations....269

8.5 Findings....270
R1: What persuasive techniques have been used by involved stakeholders to frame the SIB funding model as a 'legitimate' reform to pursue over time?....270
R2: What discourses can be identified in the highly shared tweets; and to what extent have particular discourses dominated/others been drowned out?....295

8.6 Discussion....305

8.7 Conclusion....314

Chapter 9: Discussion and concluding thoughts....318

9.1 Introduction....318

9.2 Overview of findings in relation to key research questions....320
9.2.1 What are the effects of any intervention funded by a SIB, compared to the absence of intervention or to alternative interventions, on any social outcomes?....321
9.2.2 What are the perceptions of SIB practitioners' regarding motivations for adopting SIBs; and what does this reveal about instrumental and/or institutional drivers?....322
9.2.3 How have sentiments regarding SIBs changed over time on Twitter, and did they reflect emerging (inconclusive) research findings?....323
9.2.4 Which stakeholders played a central role in communicating and promoting the SIB idea?....323
9.2.5 What discourses can be identified in the highly shared tweets; and to what extent have particular discourses dominated/others been drowned out?....324
9.2.6 What persuasive techniques have been used by SIB adopters to frame the SIB funding model as a 'legitimate' reform to pursue, over time?....325

9.3 How have efficiency and legitimacy evolved as drivers of innovation as the level of innovation efficacy has reduced?....326

9.4 What explains the truncated reform diffusion of widely but thinly diffused innovations? .327

9.5 Widely but thinly diffused ideas can still play an important role even if not fully institutionalized.....330

9.6 Theoretical reflections 331

9.7 What are the implications for evidence-based policymaking? 333

9.8 Limitations.....335

9.9 Conclusion 336

***References*.....338**

***Appendices*.....370**

Appendix A: Example Search Terms 370

Appendix B: Grey literature sources.....372

Appendix C: Hand searching of journals 375

Appendix D: Experts contacted and sample email.....377

Appendix E: Quality appraisal – effectiveness review.....380

Appendix F: Data Extraction Form – Qualitative Systematic Review 382

Appendix G: Characteristics of included studies – Effectiveness Review 385

Appendix H: Risk of Bias in Studies Referred to in effectiveness review 391

Appendix I: Quality Appraisal using CASP tool ("methodological limitations" section of CERQual Assessment) 393

Appendix J: Contribution of individual studies on the identified themes.....398

Appendix K: CERqual Assessment Scores (qualitative systematic review).....401

Appendix L: Sensitivity analysis (qualitative systematic review).....406

Appendix M: Sensitivity Analysis - Social Media Analysis (Random Sample).....409

List of tables

Table 3.1: Key Dimensions of Legitimacy

Table 4.1: Contribution of Each Empirical Component

Table 5.1: PICO Research Question

Table 5.2: Inclusion and Exclusion Criteria – Effectiveness Review

Table 6.1: SPIDER Research Question

Table 6.1: Characteristics of Included Studies – Thematic Synthesis

Table 6.2: Inclusion and Exclusion Criteria – Qualitative Review

Table 6.2*:* Factors Associated with the Adoption of SIBs

Table 6.3*:* Sampling Types Used for Qualitative Analysis

Table 6.4: GRADE-CERQual Assessment Components

Table 6.5: Characteristics of included studies – by domain

Table 6.6: Characteristics of Included Studies – Thematic Synthesis

Table 6.7: Factors Associated with the Adoption of SIBs

Table 6.8: Efficiency and Legitimacy Factors Over Time

Table 7.1: R packages Used for Sentiment Analysis

Table 7.2: Summary of Key Centrality Measures to be Considered

Table 7.3: Sample Tweets Representing Different Sentiment Levels - From Manual Coding

Table 7.4: Sample Tweets Representing Different Sentiment Levels - Based on Pretrained Models (examples of accurate coding)

Table 7.5: Comparison of Positive Sentiments (by year) - Tweets, Reddit Posts, Academic Papers

Table 7.6: Top 10 Users According to In-Degree Score

Table 7.7: Top 10 Users According to Out-Degree Score

Table 7.8: Top 10 Users According to Betweenness Score

Table 7.9: Top 10 Users According to Eigenvector Score

Table 8.1: Coding Framework – Persuasive Techniques

Table 8.2: Scale for Conflict of Interest

Levels

Table 8.3: Illustrative Quotes – Strategies of Persuasion used in Tweets

Table 8.4: Descriptive Statistics – Random Sample

Table 8.5: Types of Evidence Mentioned in Highly Retweeted Tweets

Table 8.6: Types of Events Mentioned in Highly Retweeted Tweets

Table 8.7: Exemplar Quotations

Table 8.8: Top 10 Most Highly Retweeted Tweets

Table 8.9: Top 10 Most Upvoted Reddit Posts

Table 8.10: Who Tweeted and Retweeted the 10 Most Highly Retweeted Tweets

Table 8.11: Example of a Highly Retweeted Tweet – with a High Percentage of Directly Involved SIB Stakeholders who Retweeted it

Table 9.1: SIB Evaluation - Levels of Consideration

List of figures

Figure 1.1: Structure of Peterborough SIB
Figure 1.2: SIBs Usage Since 2010 (By Year)
Figure 1.3: SIBs Usage Since 2010 (Cumulative Figures)
Figure 1.4: Policy Areas – Existing SIBs
Figure 2.1: Example of S-shaped curve (cumulative adopters) and bell-shaped curve (mean adopters per year)
Figure 2.2: Diffusion in a policy context
Figure 2.3: Model of Stages in the Innovation-Decision Process
Figure 2.4: Strategic Choices of Organizations
Figure 2.5: Nine Possible Permutations of Better, Worse, and no Change Outcomes
Figure 3.1: Different Rates of Adoption
Figure 3.2: Sources of Legitimacy – Organizational Survival
Figure 3.3: Possible Risks Introduced by Efficiency-Legitimacy Trade-off
Figure 4.1: Adoption Motivations Over Time – Classic Two Stage Model
Figure 4.2: Example of an Alternate Diffusion Model
Figure 4.3: Flow Chart for Risk Assessment - Publishing Verbatim Quotes
Figure 5.1: PRISMA Flow Chart of Study Screening
Figure 5.2: Evidence base for each intervention type
Figure 5.3: Number and types of outcomes
Figure 5.4: Chronology of evidence
Figure 5.5: Chronology of evidence (comparison of randomized and non randomized results)
Figure 5.6: Harvest Plot Summarizing Effects
Figure 6.1: PRISMA Diagram – Thematic Synthesis
Figure 7.1: SIBs Citation Tracking Data
Figure 7.2: Google Books Books Ngram Viewer Data
Figure 7.3: Twitter interest across time
Figure 7.4: Pageview Traffic Data for "Social Impact Bond (en)" Wikipedia ("Behavioural Insights Team")
Figure 7.5: Google Trends Data for the Search Term "Social Impact Bond" (term search)
Figure 7.6: Google Trends Data for the Search Topic "Social Impact Bond" (topic search)
Figure 7.7: Google Trends Views ("Behavioural Insights") & Wikipedia Page Views ("Behavioural Insights Team")

Figure 7.8: Maps showing location of tweets containing #socialimpactbond mention over time
Figure 7.9: Maps showing the spread of the SIB idea (i.e. launch of SIB) over time
Figure 7.10: Social Media Data Analysis Framework
Figure 7.11: Average Number of Tweets Containing Each Sentiment Type
Figure 7.12: Percentage of Positive, Neutral & Negative Sentiments – for Each Package
Figure 7.13: Visual Analysis of the Changing Sentiments Over Time (Positive Sentiments)
Figure 7.14: Visual Analysis of the Changing Sentiments Over Time (Neutral Sentiments)
Figure 7.15: Visual Analysis of the Changing Sentiments Over Time (Negative Sentiments)
Figure 7.16: Sentiments Contained in Original Tweets v Retweets
Figure 7.17: Comparison of Sentiments Contained in Original Tweets v Retweets – Over Time
Figure 7.18: Comparison of Sentiments Among Adopter v Non Adopter Countries
Figure 7.19: Comparison of Sentiments – Pre v Post Adoption in Adopting Countries
Figure 7.20: Sentiments Expressed in Academic Literature Over Time
Figure 7.21: Comparison to Sentiments Found in Reddit Posts From 2010 - 2020
Figure 8.1: Coding Framework – SIB Discourses
Figure 8.2: Organizational Stakeholder Types
Figure 8.3: Individual Stakeholder Types
Figure 8.4: Stakeholder Types for Adopters
Figure 8.5: Persuasive Techniques Used by Type of Stakeholder
Figure 8.6: Use of Persuasive Techniques Over Time
Figure 8.7: Use of Twitter Persuasive Techniques
Figure 8.8: Mentions of Evaluation and Alternatives – Over Time
Figure 8.9: Mentions of Evaluation and Alternatives – by Stakeholder Type
Figure 8.10: Discourses Mentioned – by Stakeholder Type
Figure 8.11: Discourses Mentioned – Over Time
Figure 8.12: Sentiments by Stakeholder Type (Whole Time Period)
Figure 8.13: Sentiments Over Time
Figure 8.14: Sentiment Category by COI Level
Figure 8.15: Most Prominent Discourse Categories in the Highly Retweeted Tweets (% of Highly Retweeted Tweets Mentioned in)
Figure 8.16: Most Prominent Sub-Categories in the Highly Retweeted Tweets (% of Highly Retweeted Tweets Mentioned in)

Figure 8.17: Percentage of Retweets Posted by Each Stakeholder Type

List of abbreviations

API: Application Programming Interface

ASSIA: Applied Social Sciences Index and Abstracts

CASP: Critical Appraisal Skills Programme

CCG: Clinical Commissioning Group

CI: Confidence Interval

COI: Conflict of Interest

COREQ: COnsolidated criteria for REporting Qualitative research

CUREC: Central University Research Ethics Committee

DIB: Development Impact Bond

ED: Emergency Department

EBM: Evidence-Based-Medicine

EBP: Evidence-Based-Practice

EBPM: Evidence-Based Policy Making

EPHPP: Effective Public Health Practice Project

EPOC: Effective Practice and Organisation of Care

EPPI-Centre: Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)

FACS: NSW Department of Family and Community Services

GRADE-CERQual: Confidence in the Evidence from Reviews of Qualitative research

GLA: Greater London Authority

GFC: Global Financial Crisis

IAAM Adoption Bond: It's All About Me Adoption Bond

IBSS: International Bibliography of the Social Sciences

IFM: Investment Fund Manager

INDIGO: International Network for Data on Impact and Government Outcomes

IRR: Incidence Rate Ratio

GO Lab: Government Outcomes Lab

MRT: Moral Reconation Therapy

MST: Multi Systemic Therapy

NEET: Not in Education, Training or Employment

NFPS: Not-For-Profit organizations

NGO: Non-governmental Organization

NHS: National Health Service

NICE: National Institute for Health and Care Excellence

NPG: New Public Governance

NPM: New Public Management

NSW: New South Wales

NQF: National Qualifications Framework

NYC: New York City

NWAU: National Weighted Activity Unit

OBC: Outcomes Based Commissioning

OHP: Out of Home Placement

ON TRACC: Transition Reintegration and Community Connection

OOHC: Out-of-Home Care

PbR: Payment by Results

PFS: Pay for Success

PICO: Population, Intervention, Comparator and Outcomes

PPP: Public Private Partnership

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PSH: Permanent Supportive Housing

PSM: Propensity Score Matching

RCT: Randomized Controlled Trial

SARA: Safety and Risk Assessment

SBB: Social Benefit Bond

SIB: Social Impact Bond

SNS: Social Networking Site

SOC: Social Outcomes Contract

SPIDER: Sample, Phenomenon of Interest, Design, Evaluation, Research type

SPV: Special Purpose Vehicle

SDG: Sustainable Development Goal

TBS: The Benevolent Society

UNDP: United Nations Development Programme

WtW: Ways to Wellnes

Part One: Background

Chapter 1: Introduction

In the context of public sector innovation, it is crucial to understand how and why some early-stage innovations rapidly diffuse through the public sector, potentially becoming institutionalized (i.e. holding 'rule-like status', with alternatives often being "literally unthinkable"), despite limited evidence of success (Zucker, 1983, p. 5). Stakeholders such as politicians, policymakers, and academics frequently claim to aspire to achieve evidence-informed policymaking. Regardless, the most technically suitable innovations are not necessarily the ones that spread and become institutionalized. In deciding whether or not to pursue an innovation, organizations care about both efficiency and legitimacy gains. Legitimacy carries many benefits; for instance, it assists with securing resources, obtaining buy-in from stakeholders and gaining autonomy from regulators. Consequently, there are many examples of legitimate (i.e. socially 'acceptable') reforms being implemented despite being inefficient; and, though harder to observe, of other, efficient reforms remaining unimplemented because they contradict external expectations. In these instances, the selected reforms could have a neutral effect with a high opportunity cost or even result in harm, making it essential to understand how and why they spread.

In exploring the diffusion of early-stage public sector innovations, I am particularly interested in exploring the 'inbetweener' category of innovation; namely, ideas that are neither widespread nor un-spread. Widespread innovations are frequently prioritized in the literature, as there is a *"massive selection bias towards ideas that ultimately become policy"* (Mehta, 2011, p. 31). It is easier to assess the diffusion of ideas that have shaped or are currently shaping institutions and policy (Smith, 2013a). Conversely, un-spread innovations are difficult to study (e.g. due to a lack of data) and are thus often ignored. The inbetweener category of ideas are worthy of investigation, as they may attract considerable attention and

resources before declining. There may be a considerable opportunity cost in initially pursuing the ideas, and the data limitations associated with un-spread ideas can be overcome by looking at thinly spread reforms (i.e. ideas that have *almost* become institutionalized).

Some level of uncertainty, defined as the lack of knowledge about the probabilities of the future state of events, is inherent in the initial stages of public sector innovation (Sniazhko, 2019, p. 2; Knight, 1921). However, the extent to which uncertainty remains post-adoption varies considerably for different innovations. Uncertainty is known to influence the extent to which an idea becomes institutionalized and is thus worthy of considering when trying to understand how and why a public reform has spread. Ongoing uncertainty about the true impact of a program "*even after adoption can cause complete institutionalization, possibly followed by mimetic abandonment when a competing institution is introduced"* (Rao et al., 2001, p. 521). Conversely, when an idea can be properly observed and measured, diffusion may not imply institutionalization as the 'institutionalization dynamic' allows for *"cycles of adoption, disappointment, and abandonment"* (Rao et al., 2001, p. 21). Ultimately, *"the more highly one evaluates an uncertain course of action ex ante, the more likely one is to be disappointed"* (Rao et al., 2001, p. 503). This is particularly likely when an innovation fails to deliver its promised benefits. Thus, it is not only whether the 'true value' of an innovation is known that is important, but also the extent to which results suggest that the innovation is beneficial (Abraham & Fairchild, 1991; Apodaca, 1952; Anderson, 1999; Barley & Kunda, 1991; Davis et al., 1994). Changing levels of uncertainty will be a central consideration in this thesis, particularly when trying to understand the link between evidence, legitimacy, and diffusion.

SIBs provide a timely example of an unproven public sector innovation that has rapidly diffused throughout many public sectors. A SIB is a form of PbR in which the commissioner pays for the outcomes achieved rather than services delivered. The capital required to fund the initiative is obtained from private investors who earn a financial return if the intervention is successful (Albertson et al., 2018). SIBs represent an administrative innovation, defined as making changes to organizational systems, as the funding model involves a change of governance rather than a change of technology (Naveh et al., 2006). In this thesis, I argue that SIBs are an example of an innovation that has been adopted for image-enhancing (i.e. legitimacy) reasons rather than efficiency reasons. There are many ways in which SIBs could be image-enhancing. For instance, although they are a subset of PbR (and thus similar to many past reforms), they have widely been regarded as new and innovative. Therefore, pursuing the idea could be beneficial, regardless of results, as *"the very fact that reform is being attempted tells the outside world that the organization is open to change and renewal"* (Brunsson & Olsen, 1993, p. 10). Additionally, SIBs draw upon private sector expertise in the governance approach and tie incentives to the delivery of outcomes. They also allegedly shift risk to the private sector, which is seen as a more 'legitimate approach' than using taxpayer funds to test new interventions.

While SIBs were likely regarded as image-enhancing when they first emerged (and for a period thereafter), I argue that the legitimacy benefits have decreased over time. Despite initial enthusiasm, SIBs usage has continued to decline in recent years, and it is unclear if SIBs have (or will) become properly institutionalized. As of 2023, there are 276 SIBs in 38 countries, with approximately $745 million USD having been raised in upfront capital (GO Lab, 2023a). The spread of SIBs has been wide and thin (i.e., to many places and sectors) rather than narrow and deep (i.e., institutionalized in particular places and sectors and sustained over time). SIBs have been launched in many jurisdictions, but there are usually

few of them in any given place (Broom, 2021). In this way, they appear to be similar to other thinly spread social innovations, such as delivery units and nudge units (which have fallen in and out of fashion), rather than thickly diffused innovations, such as shared services reforms (Gold, 2017). Given the pro-innovation bias surrounding the diffusion of innovations, it is essential to consider why SIBs have not become fully institutionalized and whether they are at all likely to (Abrahamson, 1991).

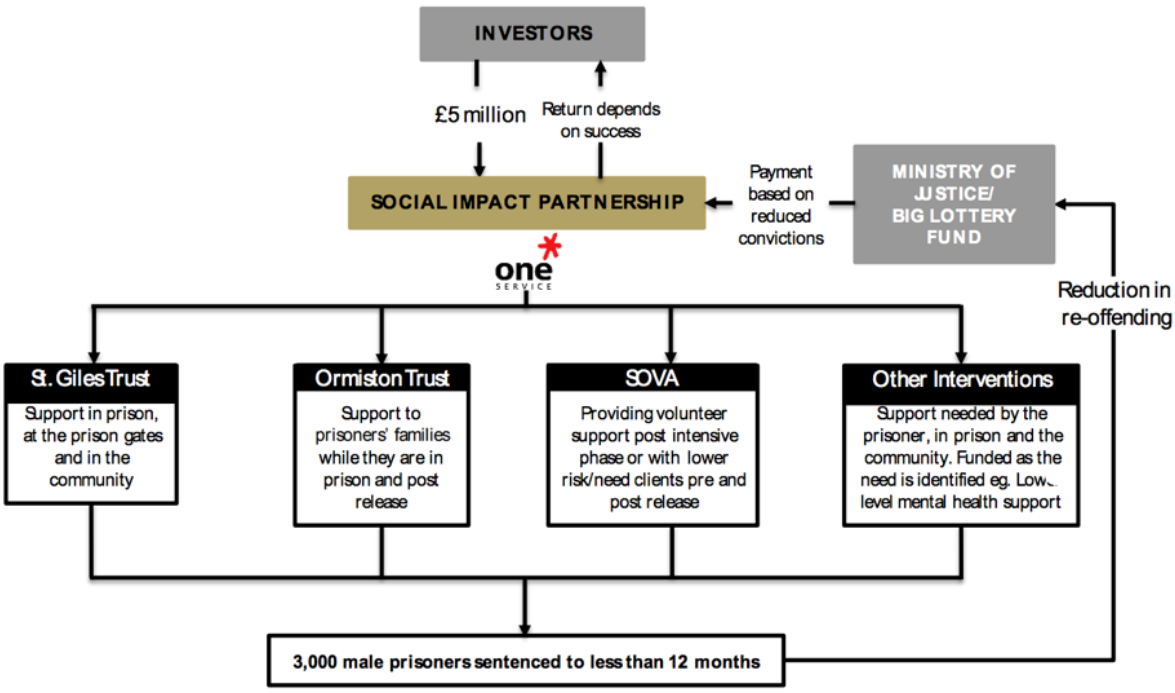
This thesis uses an organizational (sociological) institutionalist perspective to analyze what explains the rapid and widespread growth of SIBs. Unlike other theoretical lenses (e.g. rational choice or game theory), institutional theory recognizes that many institutional forms or procedures are adopted by modern organizations for legitimacy and survival reasons rather than solely for efficiency gains. I contrast the institutional view with an instrumental view, which frames organizations as rational actors who adopt institutional forms and/or procedures for efficiency reasons (Thompson, 1967; Blau & Shoenherr, 1971). Overall, this thesis assists with understanding the extent to which legitimation is central to the diffusion of public sector innovation; and the practical implications of this (e.g. whether legitimacy benefits cause decision-makers to interrogate reform ideas with lower thresholds of rigour than they otherwise would).

1.1 The emergence of SIBs

SIBs have spread throughout various public sectors since 2010, despite limited evidence of effectiveness, thus providing a clear paradox to analyze (Fraser et al., 2016; Edmiston & Nicholls, 2018; Sinclair et al., 2019). In the May 2010 General Election, the UK opposition Conservative Party promoted SIBs as a way to achieve value for money and innovation (Conservative Party, 2010). They announced the establishment of a social

investment bank to invest in innovative funding products such as SIBs. The first SIB was launched in Peterborough (Eastern England) in September 2010 (Figure 1.1). It aimed to reduce reoffending by short-sentenced prisoners through a rehabilitation intervention (Disley et al., 2011; Strickland, 2010).

Figure 1.1: Structure of Peterborough SIB

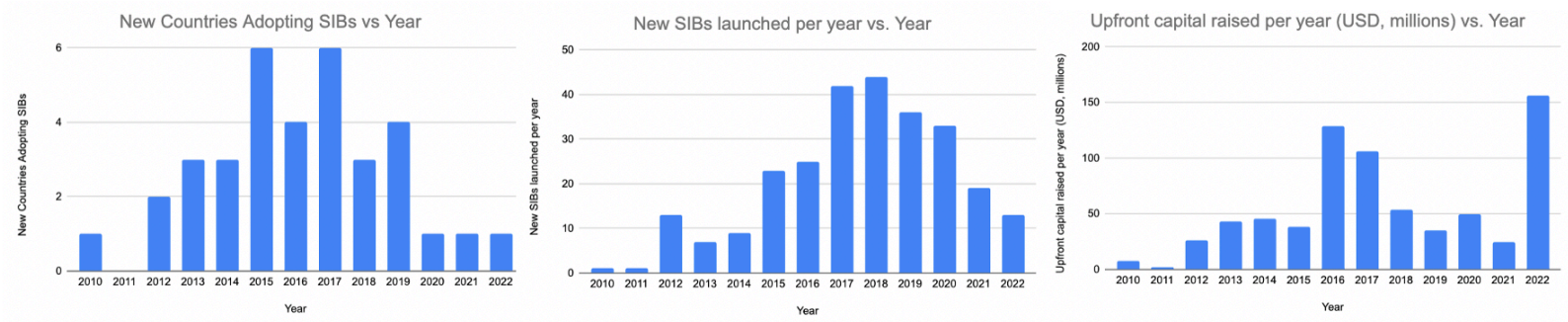


Source: Nicholls & Tomkinson (2015)

Substantial attention and capital have been directed towards the development of SIBs since then (Ronicle et al., 2014). While some stakeholders have critiqued the SIB funding model (e.g. McHugh et al., 2013; Sinclair et al., 2014; Roy et al., 2017), political enthusiasm has remained in many countries. Former British Prime Minister Boris Johnson strongly endorsed SIBs as recently as June 2022 (Patton, 2022), while Conservative MP Danny Kruger argued, *"the government is on the hunt for policies and systems that save the taxpayer*

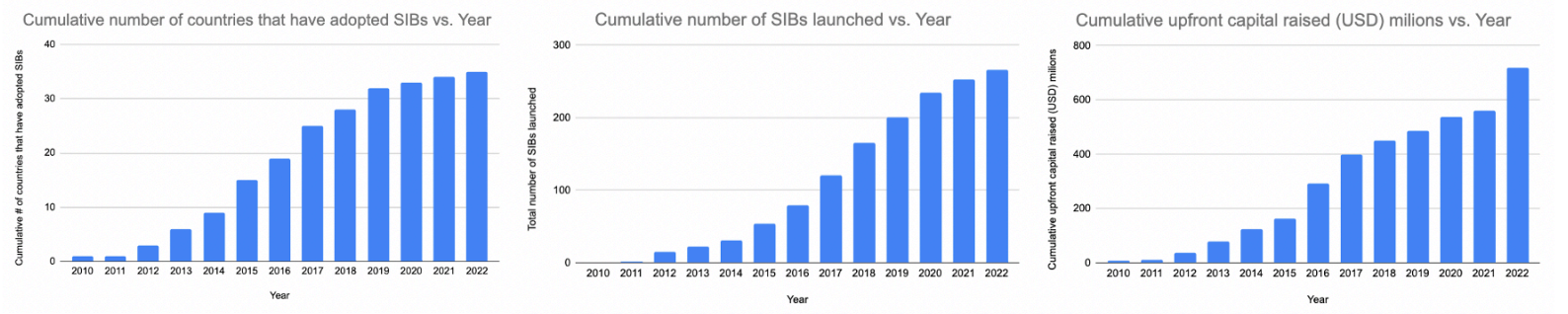
money and deliver outcomes in a quantifiable and efficient way. And this does that" (Post Reporters, 2022). In addition to 276 launched SIBs (69 of which have been completed) (Figures 7.8 and 7.9), a further 48 SIBs are in development (GO Lab, 2023b). The number of organizations involved in SIBs has grown significantly since 2010 (1 project involving 17 organizations in 2010; 266 projects involving 1,283 organizations in 2022) (GO Lab, 2023c).

Figure 1.2: SIBs Usage Since 2010 (By Year)



Source: GO Lab Data, 2023a[1]

Figure 1.3: SIBs Usage Since 2010 (Cumulative Figures)



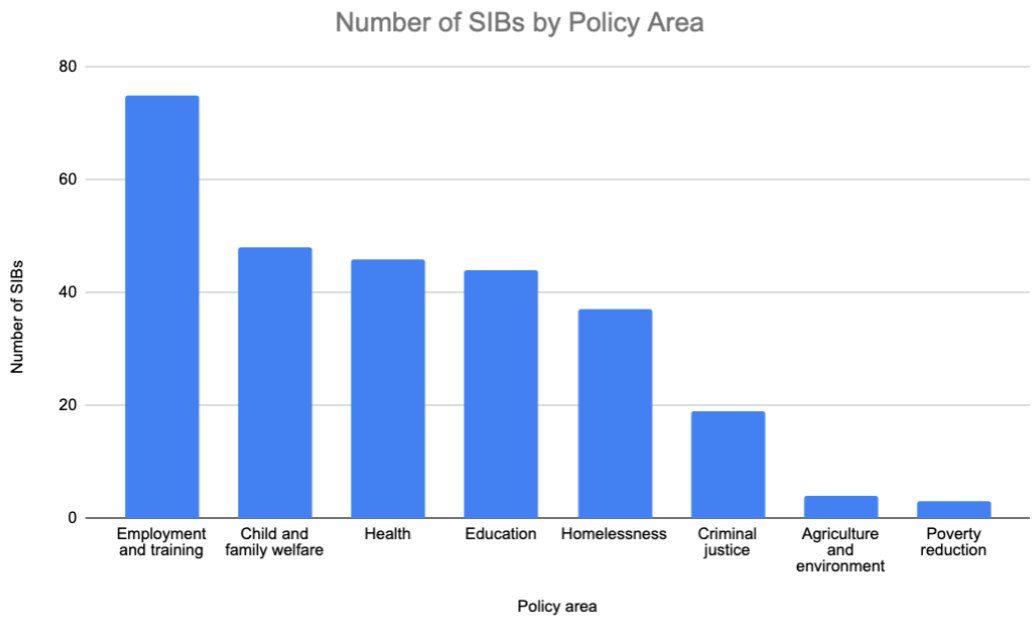
Source: Author analysis of GO Lab (2023a) data

SIBs have primarily been used in high-income countries, although, latterly, "Development Impact Bonds" (DIBs) have also emerged (UNDP, 2016; Vecchi & Casalini,

[1] Data on the number of beneficiaries engaged have not been included, as it is not possible to calculate users per year using the currently available data.

2019). DIBs typically use an external donor (e.g. aid agency, multilateral body, or a philanthropic organization) to fund outcome payments. SIBs have been used to address various policy issues, as depicted in Figure 1.4. They have grown particularly rapidly in social services that target employment, likely because this is a more straightforward outcome to observe (Hevenstone & von Bergen, 2020).

Figure 1.4: Policy Areas – Existing SIBs



Source: Author analysis of GO Lab (2023) data

The SIBs market is still in a development phase, as evidenced by the lack of a standardized design or implementation approach (Arena et al., 2016; Vecchi & Casalini, 2019). Measurement practices and financial arrangements vary greatly between contracts (Vecchi & Casalini., 2019). Rates of return are variable, there is limited evidence of how these returns have been calculated in past transactions, and the evidence base remains underdeveloped (Loxley & Hajer, 2019; Vecchi & Casalini, 2019).

1.2 A truncated diffusion trajectory

The SIB funding model rapidly appeared in successive jurisdictions but barely penetrated most of those public sectors. This could be because the legitimacy stemming from SIBs has partly come from the novelty surrounding the idea. Thus, the spread of the idea has likely been truncated by diminishing levels of novelty over time. The shift to DIBs (potentially to recreate the novelty factor in a new policy domain) further truncated the trajectory of SIBs. Reforms are often presented as 'new' to avoid *"disbelief on the grounds that the same doctrines have been tried before, shown to be unsatisfactory in particular cases, and abandoned"* (Hood & Jackson 1991, p. 194). However, in the case of SIBs shifting to DIBs, the 'new' DIB idea has explicitly acknowledged its lineage (e.g. with an almost identical name). Thus, I argue that DIBs have been an attempt to prolong the SIBs idea, rather than a complete reframing of the idea.

Another possible explanation for the truncated diffusion is that the funding model represents a 'chameleonic idea' (Smith, 2013) or 'vehicular' idea, *"which [is] imbued with an 'ineliminable vagueness and "mobility" that allows [it] to transform with relative ease as they move between actors and across contexts"* (Carroll, 1994; McLennan, 2004, p. 485). Elasticity and transformability are key characteristics which enable ideas to travel between actors and across contexts. Fraser et al., (2020, p. 196) note that *"the idea of what a SIB is, or should be, is imbued with 'chameleonic' characteristics".* Thus, even if the SIB concept is not institutionalized, it represents an idea that is elastic enough to fit within accepted institutional boundaries. Smith (2013, p. 192) holds that the *"concept of chameleonic ideas can be used to help explain both the 'fractured journeys' of psychosocial and lifecourse theories and the 'partial journey' of material-structural ideas"*. It could, therefore, partly or

wholly explain why SIBs have easily travelled to new contexts, enabling proponents to benefit from prolonged uncertainty and a renewed sense of innovativeness.

1.3 Thesis outline

This thesis uses a monograph format and consists of three parts, all aimed at addressing the following overarching research question: *"what explains the rapid and widespread growth of SIBs?"* Each of these sections are briefly discussed below.

Background

This section contains four chapters that situate the study in the broader literature on public sector innovation and explain the research design. Chapter 1 introduces the problem to be analyzed and presents SIBs as the specific innovation to be studied. I argue that SIBs provide an example of an idea spreading rapidly despite limited evidence of success, partly or wholly for image-enhancing reasons. Chapter 2 provides a literature review focused on understanding what is known about the diffusion of early-stage public sector innovation. I present the review in three sections: 1) definition, types and stages of innovation; 2) drivers of innovation; and 3) outcomes of innovation. I highlight how uncertainty is a persistent and vital theme when considering why SIBs have spread, as it influences various drivers of adoption (e.g. mimicry) and allows for rhetoric to replace concrete results and evidence. Chapter 3 outlines the theoretical framework to be used for this thesis. I present the instrumental and institutional views and argue that the latter (particularly legitimacy theory) is required for this thesis as SIBs provide an example of organizations trying to enhance their image by adopting the idea. Finally, Chapter 4 outlines the methodology for the empirical sections. I focus on the overarching study design and how it fits together through a 'consilience' framework, which involves combining evidence from multiple sources to reach

inferences from valuable but individually incomplete data (Hood & Dixon, 2010; Hood & Dixon, 2015).

Empirical research

This section consists of four chapters, each of which help to address the overarching research question, step-by-step. Firstly, Chapter 5, consists of an effectiveness systematic review conducted to understand how uncertainty regarding intervention effectiveness has changed over time. I highlight a severe lack of quantitative evidence on SIB effectiveness, across the entire diffusion period, with robust (program-level) quantitative results only emerging quite recently, and results about the impact of the funding model still lacking. Rather than leading to the early abandonment of the idea, this has assisted with policy diffusion, since it has not been possible to prove that the idea, with its powerful rhetorical appeal, does not in fact work in practice. Chapter 6 builds on this with a qualitative systematic review. Using a thematic synthesis approach, I consider how the themes identified relate to the well-documented pillars of legitimacy (Binz et al., 2016; Suchman, 1995; Ruef & Scott, 1998; Walker et al., 2014; Thomas & Harden, 2008). I find strong evidence for various legitimacy pillars (particularly normative and pragmatic), suggesting that concern among governments to secure legitimacy among valued stakeholders by adopting a novel idea with strong private-sector overtones has driven the rapid and widespread adoption of SIBs.

As there have been few attempts to synthesize the complete landscape of SIBs literature, the systematic reviews provide a timely and important contribution. In isolation, however, the systematic reviews do not provide conclusive answers as to what explains the rapid and widespread growth of SIBs. Other forms of data are thus required, particularly

previously overlooked forms of data. Recognizing this, Chapter 7 presents a quantitative analysis of an original Twitter dataset, consisting of 13,836 individual tweets collected over more than 10 years (2010 – 2020). A combination of statistical analysis, sentiment analysis and network analysis are used to understand how levels of interest and attitudes towards SIBs have changed over time and to identify which stakeholders have been particularly influential in spreading the idea online. I find that while annual academic citations have continued to grow, online interest has declined; noting that SIB usage has also declined, this does not appear to be due to taken-for-grantedness (i.e. the idea being institutionalized). I also find that negative sentiments have largely been drowned out across the entire time-period, with the dataset featuring overwhelmingly positive tweets. Additionally, as predicted by the existing literature, I find strong evidence of "fashion setting" stakeholders, such as consultants, playing a very active and influential role in the spread of the idea, particularly in the early stages of the management fashion life cycle (Heusinkveld, 2004).

In Chapter 8, I conduct framework analysis to understand what pre-existing frameworks from the public administration and SIBs literature reveal about how and why SIBs have thinly diffused. I apply a framework developed in a prior SIB-focused discourse analysis (Ormiston et al., 2020) to a purposive sample of the Twitter data to identify which discourses have been prevalent and shared widely and which have been drowned out (i.e. shared less frequently compared to other discourses). The ‘social finance’ and ‘impact’ discourses are well-represented, with less of a focus on the ‘collaboration’ discourse. I also identify several cautionary discourses related to possible harm (e.g., marketization of vulnerable people). I find that robust academic studies and impact evaluations are rarely mentioned in the top retweets, and that the ‘events’ mentioned suggest that policymakers often establish SIBs with great enthusiasm yet fail to properly evaluate the results. Finally, I

apply a well-accepted framework of public administrative persuasive strategies (Hood & Jackson, 1991) to a random sample of tweets to identify how the value of SIBs has been constructed and conveyed. I find that strategies of persuasion have been used across time, even those that appear to be time-sensitive (e.g., framing the idea as new). The absence of evidence (identified in Chapter 5) has allowed for these persuasive techniques to take hold, with SIBs spreading based off a rhetorical process, rather than strong evidence of success. I also devise and apply a conflict of interest (COI) scale and identify that actors who benefit from the success of the idea (i.e., higher COI) have expressed particularly positive sentiments and played a crucial role in propelling the diffusion of the idea online.

Interpretations and Contributions

This final section consists of a stand-alone chapter in which I identify common themes and overarching lessons. I reflect on how the differing data sources and methods, drawn together using a consilience approach, have contributed to the overarching research question. The consilience approach allowed for a novel comparison of efficiency and legitimacy motivations, which built on the approaches typically used to understand how and why management reforms have spread (Hood & Dixon, 2010; Whewell, 1840). In this chapter, I re-state the overall findings of each phase of the DPhil project while identifying common themes that cut across multiple chapters. As the empirical components of the thesis uncover evidence of institutional motives influencing the uptake of SIBs, I argue that legitimation is central to the diffusion and potential institutionalization of public sector innovation. However, legitimacy alone can only propel an idea for so long, leading to the eventual demise of some reforms, such as SIBs. Finally, this Chapter considers what the findings reveal about the notion of ‘evidence-based’ policymaking and how governments decide on which public sector reforms to pursue. In sum, the findings of this thesis will help

decision makers (e.g., policy makers) to make better-informed decisions about the adoption (or otherwise) of early-stage, unproven public sector reforms.

1.4 Conclusion

In combination, the analyses presented in this thesis generate important insights regarding the link between legitimacy, evidence of performance and diffusion of innovations in the public sector. Noting the need for a more flexible model of diffusion that accounts for ongoing uncertainty, I focus on how motivations have changed over time, as the level of uncertainty regarding intervention effectiveness has changed. Methodologically, this thesis demonstrates the limitations of past approaches in explaining the rapid growth of some public sector innovations (e.g. citation tracking, regression models using city level characteristics etc); and explores how emerging forms of data, particularly social media data, can be incorporated into empirical studies. More practically, the empirical analyses generate a series of higher-level findings (presented in Chapter 9), which are relevant to policymakers and other stakeholders who are involved with deciding which early-stage public sector innovations to pursue and which to overlook.

In each of the following chapters, I will reflect on how and why SIBs, a widely but thinly diffused innovation, have continued to spread, absent evidence of effectiveness. I will demonstrate how strong institutional motives can lead to a truncated diffusion trajectory, as legitimacy benefits may eventually erode, in the absence of evidence, and domain switching may be used, to reframe the idea as a legitimate one to pursue, to a new set of stakeholders. In each chapter, I will make various links to the concept of uncertainty, for instance, in noting that ongoing uncertainty about the effectiveness (or otherwise) of an idea, creates space for rhetoric to take hold and for ideas to spread based on persuasiveness rather than results. A

lack of certainty about the effectiveness of an idea assists with diffusion, potentially helping to explain why evidence is sometimes overlooked and/or ignored. While they may not become fully institutionalized, widely but thinly diffused innovations such as SIBs may continue to spread if it is not possible to conclude that they do not work. Although an appealing and persuasive idea, I demonstrate that SIBs have been pursued at a considerable opportunity cost, as substantial resources have been devoted to the continued spread of the idea and other, more efficient reforms, may have been overlooked.

Chapter 2: Literature Review

2.1 Introduction

In this chapter, I present a literature review focused on the diffusion of early-stage public sector innovation. I focus on this subset of the literature as uncertainty is greatest in the initial stages of innovation, and my research question focuses on understanding why specific early-stage public sector innovations, for which there is often minimal evidence of success (and thus high levels of uncertainty), 'take off'. I structure the literature review around the following three sections: 1) definition, types and stages of innovation; 2) drivers of innovation; and 3) outcomes of innovation. In exploring what is already known about these factors, I develop the motivation and conceptual foundations for this thesis. Within the literature focused on the diffusion of institutional innovation, strong interest remains in patterns of diffusion. However, interest has shifted from simply showing patterns of diffusion (considered to be indicators of institutional effects), to other questions, such as the extent to which organizations adopt similar forms and structures due to adapting to *"what they believe society expects of them"* (Greenwood et al., 2017, p. 6). There is also interest in examining what happens after adoption, to the practices or organizations that have diffused (e.g. Ansari et al., 2010; Weber et al., 2009; Shipilov et al., 2010). In this chapter, I take into account this shift and consider how existing research on the diffusion of public sector innovation can help to explain the rapid and widespread growth of SIBs.

2.2 Definition, types and stages of innovation

Various definitions have emerged for the term 'innovation.' In his frequently quoted definition, Rogers (1995, p. 11) argues notes that *"an innovation is an idea, practice, or object that is perceived as new by an individual or other unit of adoption".* Thus, he argues, it does not matter whether the idea is objectively new. Although useful in the context of

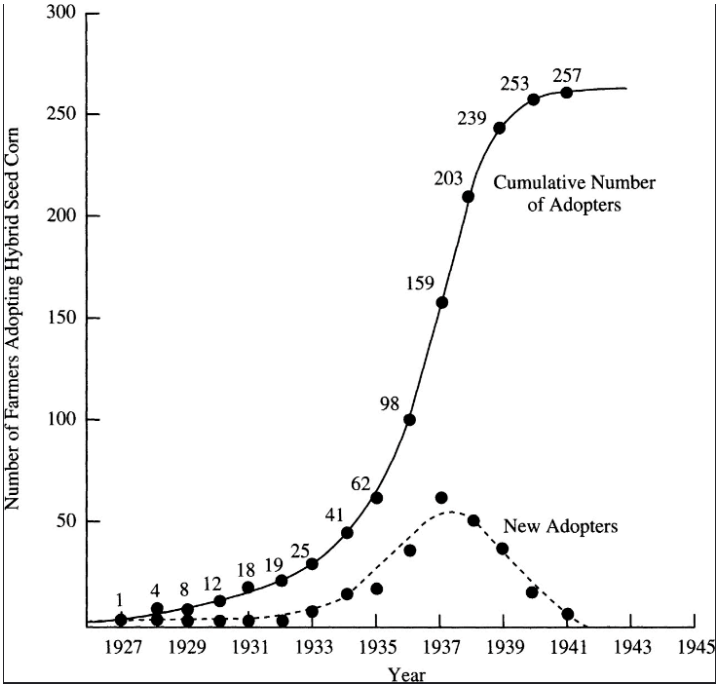
individual behaviour, this definition is less applicable to an organizational context as stakeholders in an organization typically need to do more than simply perceive an innovation as new (Greenhalgh et al., 2004). No consensus definition exists for organizational innovation', which has been *used "in a rather loose and slippery manner in many writings"* (Lam, 2006, p. 139). In this thesis, I apply Greenhalgh et al's., (2020) definition of 'health innovation', more broadly, to the context of public sector innovation (Lam, 2006). This definition notes that innovation refers to *"a set of behaviors, routines, and ways of working (along with associated technologies), which are perceived as new, linked to the management of a condition or the provision of services, discontinuous with previous practice (i.e., not just quality improvement), directed at improving outcomes for service users and/or staff, and implemented by means of planned and coordinated action"* (Greenhalgh et al., 2020, p. 2). As the definition notes that an innovation is discontinuous with previous practice, it implies that there is inherent uncertainty associated with innovating. It is integral to distinguish between the terms 'innovation' and 'reform'. Public management reforms may not necessarily be innovative, as the term refers to *"deliberate changes to the structures and processes of public sector organizations with the objective of getting them (in some sense) to run better"* (Pollitt and Bouckaert, 2011. p 2). Diffusion is the phenomenon of social influence which leads to adoption, and adoption is defined as the decision to engage with and use a particular innovation (Greenhalgh et al., 2020).

Innovation type is critical to consider as it influences the innovation adoption process. Three categories of innovation are well-recognized in the existing literature: 1) product and process innovation (Kubeczko et al., 2006); 2) radical and incremental innovation (Tushman & Anderson, 1986); and 3) technical and administrative innovation (Daft, 1978; Tanninen et al., 2008). I focus specifically on the adoption of administrative innovation. Technical

innovations focus on the products, services and/or production processes and affect the technical system; they are thus measurable, and the advantages are often reasonably clear (Daft, 1978; Damanpour et al., 1989). Conversely, administrative innovations *"involve an administrative component and affect the organization's social system"* (Tanninen et al., 2008, pp. 3-4). SIBs involve a change in governance processes and funding arrangements and thus meet the definition of administrative innovation. However, they confound administrative innovation with technical innovation, as the funding mechanism can be used to fund novel interventions. As discussed in Chapter 5, this creates evaluation challenges, as it is difficult to measure the impact of the funding mechanism itself. Consequently, even if positive outcomes are achieved, it is difficult to determine whether better outcomes could have been achieved with a different (cheaper and less complex) governance arrangement.

The diffusion of innovations literature focuses on how particular innovations become a phenomenon in reality. Thus, there is a particular emphasis on how adoption decisions are made. Innovations commonly exhibit the S-curve pattern. This pattern has been demonstrated by many empirical studies, starting with Ryan and Gross (1943), who found that the number of hybrid seed corn adopters per year approached an S-shaped curve over time; conversely, the frequency distribution of the mean number of adopters per year approached a normal, bell-shaped curve (Figure 2.1). The initial adoption of change is typically slow, involving experimentation and 'trial and error'. When a few other actors have adopted the innovation, there is often a steep climb in adoption rates before a levelling-off. When institutional change has reached the levelling off stage, further investment to gain additional adopters is usually 'wasted' (Frederickson et al., 2016). In this thesis, I am particularly interested in the early stages of the adoption curve, where 'innovators' and 'early adopters' are pursuing the idea, despite high levels of uncertainty (Rogers, 1995).

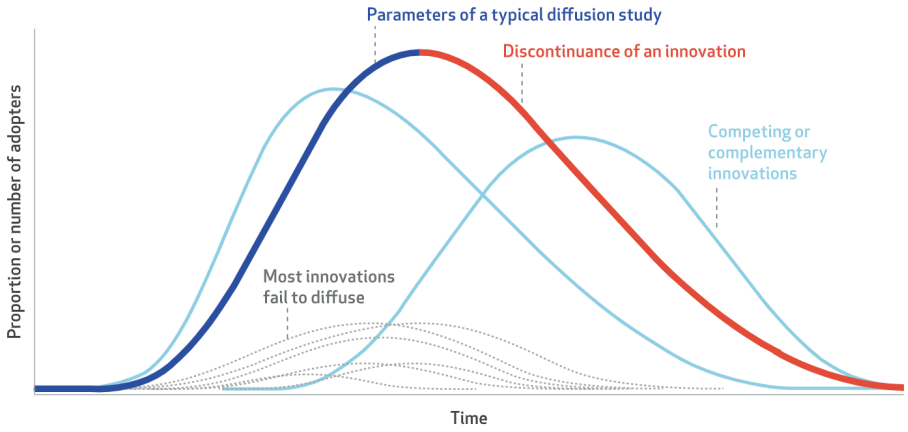
Figure 2.1: Example of S-shaped curve (cumulative adopters) and bell-shaped curve (mean adopters per year)



Source: Rogers (2003, p. 281), based on Ryan and Gross (1943)

The S-shaped curve only describes cases of successful innovation (i.e. where an innovation has spread to almost all of the potential adopters in a system), and thus may not be fully applicable to the context of SIBs, in which the diffusion has been truncated and barely penetrated many public sectors. Dearing and Cox (2018, p. 184) produce a model (Figure 2.2) that is more applicable to a policy context. Their model *"illustrates the relationships between rates of adoption and how we characterize diffusion under different scenarios, including when innovations are introduced and do not diffuse"* (Dearing & Cox 2018, p. 184). It recognizes that in the context of public sector innovation*, "time to adoption can be shorter because of the occasional convergence of national attention to a problem, financial incentives, readiness for change among elected officials, motivated and organized groups, and an innovative solution that is perceived positively"* (Dearing & Cox 2018, p. 184).

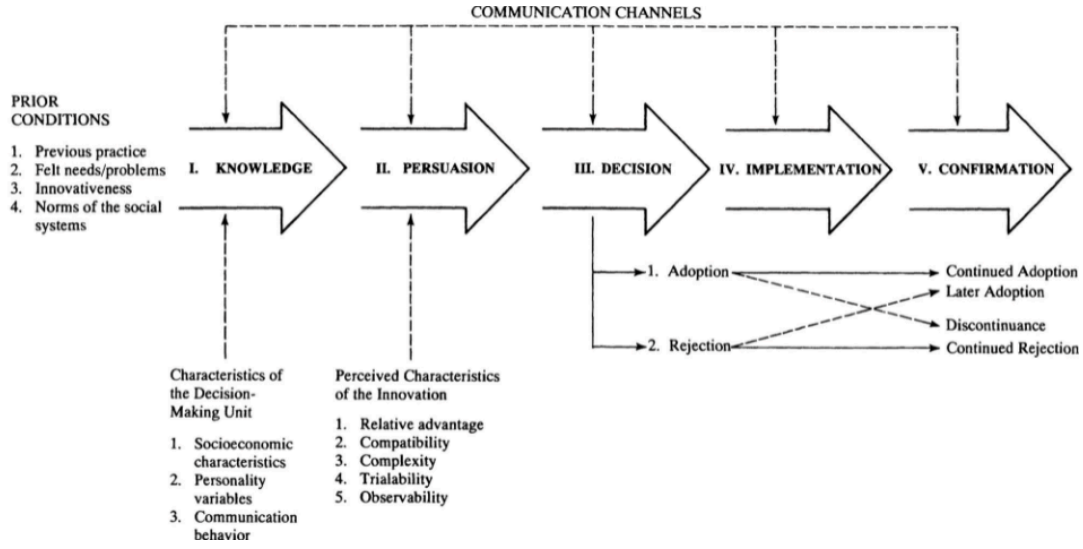
Figure 2.2: Diffusion in a policy context



Source: Dearing and Cox (2018, p. 184)

Traditionally, the individual adoption process has been presented as having five stages (Figure 2.3). These include: 1) initial knowledge of the innovation; 2) forming an attitude about the innovation; 3) making a decision to either adopt or reject the innovation; 4) implementation of the idea; and 5) confirmation of the decision (Rogers, 1995). While depicted linearly, organizational innovation is "*generally a nonlinear process characterized by multiple shocks, setbacks, and unanticipated events"* (Greenhalgh et al., 2004, p. 610; Van de Ven et al., 1999). The initial stages (up to and including the "decision stage"), are the most relevant for the context of early-stage public sector innovation.

Figure 2.3: Model of Stages in the Innovation-Decision Process



Source: Rogers (1995, p. 185)

The innovation-decision process, as formulated by Rogers (1995; 2003) has frequently been drawn upon in the empirical literature. In some instances, the data contained in empirical studies have supported Rogers' (2003) innovation-decision process. For example, Walitzer et al., (2015) explored the dissemination and adoption of an addiction outpatient clinic intervention; 141 primary administrators and 837 clinicians completed questionnaires that assessed aspects of the five innovation-decision stages. Overall, as hypothesized, they found their data to be consistent with the Rogers' (2003) diffusion model. However, while many empirical studies have supported the original model, there are several limitations. Firstly, much of the evidence used to inform the theory did not originate in the field of early-stage public sector administration. Secondly, the theory is more applicable to the adoption of behaviours rather than the abandonment or prevention of behaviours. Finally, the innovation-decision process model does not consider factors such as an individual's or organizations' resources or social support to adopt the new behaviour. Additionally, the

model should not be used to *"overgeneralize to complex, process-based innovations in service organizations, for which the unit of adoption (at this level, more often called assimilation) is the team, department, or organization in which various changes in structures or ways of working will be required"* (Greenhalgh et al., 2004, p. 600).

In some studies, alternate models have been applied and found to be more suitable. For instance, Greenhalgh et al., (2004) found that a lesser-known model (Concerns Based Adoption Model), which was developed for innovation in schools, better accounted for the findings of empirical studies of complex services innovations in an organizational context (Hall & Hord, 1987). Additionally, some studies have noted that research examining the non-adoption of innovations is limited and thus important to focus on. For example, MacVaugh and Schiavone (2010) developed and applied an integrative model of non-adoption. With these mixed findings in mind, I will build on Roger's model by considering how and where efficiency and legitimacy drivers fit into the innovation-decision process; and what happens when evidence emerges *against* one of these drivers (e.g. if the accumulated evidence means that stakeholders move from being uncertain about effectiveness to certain of ineffectiveness).

2.3 Drivers of innovation

The drivers of innovation have been discussed extensively in the existing literature, with research focused on understanding what encourages and prevents public sector innovation. Firms may innovate *"both when successful and unsuccessful"* (Cyert & March, 1992: 189; Kotiloglu et al., 2019), and widespread adoption of an idea cannot be explained by relative advantage alone (Greenhalgh et al., 2004). Even innovations that are viewed as being evidence-based are often subject to a lengthy period of negotiation amongst potential

adopters and are sometimes still reframed or dismissed (Denis et al., 2002; Ferlie et al., 2001; Fitzgerald et al., 2002; Grimshaw et al., 2004). In this section, I provide various examples as to why organizations innovate and describe why organizations may fail to innovate and/or abandon innovations.

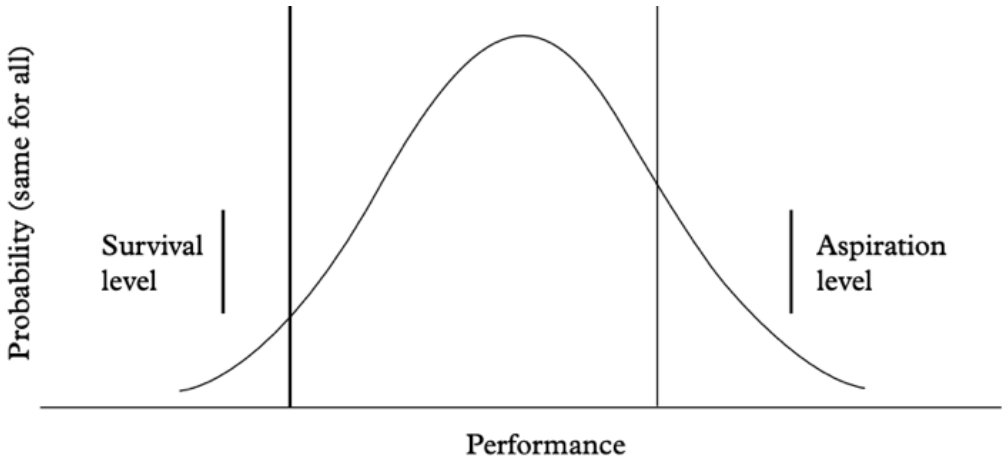
The possible explanations for pursuing an innovation range in the extent to which they follow the orthodox, rational view rather than the institutional view. Broadly, adoption reasons related to the instrumental view centre around the fact that managers identify a problem and search for a solution in response (e.g. to address a gap in performance). Critiques to this view highlight the fact that rational models of policymaking are impossible to fulfil in practice and policymakers find solutions that are 'good enough' as they are unable to investigate all policy alternatives. Drivers related to the institutional view note that the solution comes first, as proponents search for problems that the solution could solve; new institutions are thus *"solutions searching for problems"* (DiMaggio, 1988; Greve, 2003, p. 173). Other possible reasons, related to the institutional view, include: 1) reform fashions can quickly spread ideas; 2) to deal with uncertainty, organizations may mimic other organizations; 3) persuasive ideas are hard to reject, particularly when packaged in a particular way; and 4) it can be beneficial to pursue novel ideas regardless of the effectiveness. In discussing the drivers of innovation in further detail (below), I have ordered these by the most to least orthodox/rational.

2.3.1 Why organizations innovate

Problem searching for a solution

The orthodox, rational view of innovation holds that an organization experiences a problem and seeks and adopts a solution to overcome this. Under this lens, problems are thus perceived to come first. Managers are ultimately concerned with an aspiration level (the performance they seek to achieve) and a survival level (the minimum level of performance that will allow them to continue operating) (Figure 2.4) (Greve, 2003). Performance gaps prompt the diffusion of innovations amongst organizations that can efficiently close these gaps by adopting the idea. If an innovation meets an identified need or problem, an individual or organization is thus more likely to adopt it (Greenhalgh et al., 2004; Hall & Hord, 1987; Wejnert, 2002). Proponents of institutions *"act as "solution entrepreneurs" who use a problem of low performance to argue for the adoption of their favorite institution"* (Greve, 2003, p. 173). As organizations are more predisposed to look for solutions when performing poorly, those with performance levels below the 'aspiration level' "*have higher rates of strategic change, R&D expenditure, innovation, and investment"* (Greve, 2003, p. 1).

Figure 2.4: Strategic Choices of Organizations



Source: Greve (2003)

Various empirical studies have supported the notion that organizations reduce the intensity of actions (e.g. new product introductions, research and development work, strategic

changes) when performance increases above aspiration levels, as there is a reduced need for problemistic search (Alessandri & Pattit, 2014; Greve, 2003b; Kacperczyk et al., 2014; Lehman & Hahn, 2013; Lucas et al., 2018; Joseph & Gaba, 2015; Parker et al., 2017; Lin, 2014). However, others have found that organizations performing above aspiration levels increase the intensity of their actions as increased performance produces 'organizational slack' that can encourage exploration (Baum et al., 2005; Bromiley & Washburn, 2011; Iyer & Miller, 2008; Lounsbury & Beckman, 2015; Salge et al., 2015; Baum et al., 2005). Thus, empirical results relating to the effects of performance feedback often vary.

Recognizing the empirical inconsistencies, meta-analytic methods have been used to quantitatively assess the extent to which empirical evidence supports the propositions underpinning performance feedback theory. Kotiloglu, Chen and Lechler (2019) statistically analyzed 113 studies (containing 302 effect sizes) and found that performance below aspirations intensifies all forms of organizational actions (both short-term and longer-term), thus providing robust evidence for the notion that performance below aspirations triggers problemistic search (Ben-Oz & Greve, 2015; Cyert & March, 1963; Greve, 2003c). The results highlighted that performance below aspirations is more likely to intensify organizational actions than performance above aspirations is to reduce them. Thus, organizations are averse to performance below aspirations and will attempt to address performance shortfalls (Greve, 1998).

While it is well-recognized that organizational decision-makers will search for a satisfactory solution when faced with a problem (e.g. Cyert & March, 1992; van der Voet & Lems, 2022), the required solution cannot be identified through searching alone, for complex societal problems. Instead, a novel solution must be created. Empirical studies (e.g. van der

Voet & Lems, 2022) have found that negative performance decreases the creativity of policy solutions that are proposed by decision makers. This may be because decision-makers perceive negative performance to be a threatening situation, and individuals, groups and organizations are likely to behave rigidly (e.g. prioritize efficiency rather than flexibility and adaptation) in response to threatening situations (Schwarz et a., 2022; Staw et al., 1981). Similarly, Hansen and Nielsen (2022) conducted two experiments among managers working in public education in Denmark, finding that while low performance affects problem focus, it does not influence managers' intention to engage in innovative search. Additionally, low performance could be linked to enhanced preferences for performance-related changes but not for different forms and types of changes.

Finding solutions that are 'good enough'

Less rational explanations as to why ideas spread hold that rational models of policymaking are impossible to fulfil in practice due to limits to rationality faced by administrators, particularly in relation to *"memory, attention and capacity"* (Lindblom, 1968; Augier & March, 2001; Schwarz et al., 2022, p. 902; Thornton et al., 2012). The concept of 'bounded rationality' holds that individuals cannot deal with complex problems in a manner that meets the demands of objective rationality (Simon, 1955; Lindbolm, 1959). Due to their 'bounded rationality', public sector organizations face inherent uncertainty and *"increasingly, need to deal with unknowns and "Black Swans," which are consequential shocks that cannot be anticipated based on past experience"* (Schwarz et al., 2022, p. 902; Taleb, 2010). Investigating all policy alternatives is beyond the analytical competence of policymakers who rarely have the level of knowledge required to make a rational choice. As a result, policymakers often report to 'satisficing' (an amalgamation of 'satisfy' and 'suffice') and making decisions which are 'good enough' (Simon, 1956; Schwarz et al., 2022). They

'muddle through', often focusing on small-scale, incremental policy change as this enables them to reduce the number of alternatives to be considered and the complexity of the policy process (Lindblom, 1959; Lindblom & Woodhouse, 1980). Incrementalism does not prevent large-scale policy change; instead, it conceptualizes any large-scale changes as the accumulation of multiple smaller developments. Under this lens, if a solution is already known and within reach, policymakers may implement it rather than searching further afield and/or for longer.

Dynamic, 'chaotic' policy making

There have also been more radical critiques of rational policymaking frameworks, for example, Cohen, March, and Olsen's (1972) 'garbage can' model. This model accounts for the chaotic reality of policymaking by recognizing that it is a dynamic, unpredictable and often irrational process. A garbage can metaphor is used, in which policy problems and solutions are 'dumped' in a metaphorical garbage can. Policy problems and solutions can be paired when opportunities arise to make new policy decisions. Thus, *"existing proposals ... can be passed off as solutions to new problems"* (Stone, 2001, p. 11); and solutions can pre-date problems rather than following the rational analysis of problems (Nutley et al., 2007). Kingdon's (1984) 'multiple streams' model provides a further refinement to the 'garbage can' model. It focuses on understanding the agenda-setting stage of the policy process (and specifically how major issues appear on the agenda). Kingdon (1984) holds that policymaking involves three separate streams (problem stream; policy stream; and political stream); when the streams join together, the greatest policy change occurs. Specifically, policy problems are combined with solutions and enter the political stream when 'policy windows' open. Policy windows rarely open and do not remain open for long. When this

occurs, policy entrepreneurs play a crucial role in coupling solutions to problems and political opportunities (Neilson, 2001).

Reform "fashions" can spread inefficient ideas

The continuous emergence of new management practices, trends and fashions ensures that the practices of an organization will *"at least sometimes appear old-fashioned and in need of reform"* (Brunsson & Olsen 1993, p. 37). Management fashions are a subset of all the available management ideas, specifically, the concepts that *"relatively speedily gain large shares in the public management discourse"* (Jung & Kieser 201, p. 329). The terms "fad" or "fashion" are often used to refer to diffused technologies that are not necessarily efficient, have little organizational utility and/or may serve symbolic functions (e.g demonstrating innovation), but have little effect on economic performance (Abrahamson, 1986, 1991; Fombrun, 1987). However, innovations do not always 'go wrong'; fashionable ideas may encourage managers to identify problems they would otherwise overlook, disseminate "good" ideas, or use management knowledge they may not otherwise encounter.

Nonetheless, while fashions be a valuable conduit for introducing new management ideas to organizations that are otherwise full of inertia, they can also come at a high cost. Fads or fashions can be harmful as they may cause the rejection of other technologies that have the potential to become technically efficient (Abrahamson, 1991; Teece, 1977). They may prompt organizations to *"leap rapidly from one technology to the next"*, thus meaning that no technology is given enough time to demonstrate that it works (Abrahamson 1991, p. 589; Hackman, 1975; Lawler & Mohrman, 1985). Similarly, as decision-makers' jump rapidly' from problem to problem, *"these problems go out of fashion before technologies can solve them"* (Abrahamson, 1991, p. 589; March & Olsen, 1976; Schon, 1971). Fads or

fashions may facilitate the uptake of technically efficient technologies in some instances and for some organizations but not others (Rogers, 1962, 1983). Inefficient fashions may also spread because we do not know the outcome of an idea until we have tried it; it is only with hindsight that an idea may be considered a "bad" innovation.

A powerful set of actors (such as consulting firms, investing firms, management gurus) spread pre-packaged ideas and solutions, impelling the diffusion of fashionable ideas. Rather than ideas circulating widely due to being 'powerful', ideas *become* 'powerful' as they circulate, for reasons such as *"who transports or supports them"* (Czarniawska & Joerges, 1996; Greenwood et al., 2017, p. 105; Rovik, 1998). Management consultants create an initial wave of interest which can lead to a critical mass of adopters, thus triggering further 'bandwagon effects' in the management fashion market (Madsen & Slåtten, 2015, p. 4). Thus, another driver of innovation is the fact that there are organizations that are dedicated to providing solutions (management fashions) to other organizations (Brunsson & Olsen, 1993). Policy ideas are intentionally framed to increase their support and take-up (Abrahamson, 1996; Béland, 2005).

Imitation

Imitation is likely when managers face uncertainty in decision-making (Cyert & March, 1963). Social proof, defined as *"using the actions of others to infer the value of a course of action"*, is particularly influential in instances where decision-makers are uncertain about the true impart of a particular reform and when they can observe the actions of similar peers (Rao et al., 2001, p. 502). Thus, another reason for innovating is imitation, as decision-makers imitate their peers, particularly those who are perceived as successful and legitimate. Imitation carries many benefits; for instance, it minimizes search costs and avoids

experimentation costs; *"if a practice is prevalent among competent actors, then it must be sensible"* (DiMaggio & Powell, 1983; Rao et al, p. 502). However, imitation is a 'fragile' reason for adoption. Thus, to the extent that this driver has underpinned an adoption decision, *"we should expect to see overevaluation, disappointment, and abandonment"* (Rao et al., 2001, p. 503). If an innovation has been pursued due to imitation, stakeholders may intentionally overlook uncertainty or assume that there is certainty (of effectiveness), without properly investigating whether the 'evidence' is methodologically robust.

Persuasive ideas

Innovations may also be pursued because they are so compelling that it is difficult for policymakers not to adopt them. Conversely, unpersuasive ideas are often perceived as "bad ideas" which remain unimplemented; if a project is 'insufficiently appealing', it is unlikely to attract the support of key opinion leaders and spread (Locock et al., 2001; Greenhalgh et al., 2004). The superseding of one idea over another in public administration often occurs through a process that relies on rhetoric rather than concrete evidence and extensive examination of the available data (Hood & Jackson, 1991). Thus, ideas may become powerful due to "how they are packaged, formulated and timed" (Czarniawska & Joerges, 1996; Greenwood et al., 2017, p. 105; Rovik, 1998). Rhetoric, defined as "the art of using language effectively" (Smullen, 2010, p. 45), plays a vital role in the diffusion of public sector ideas. Through a social process, some ideas 'win' over potential rivals, ultimately being labelled as the preferred view and afforded prominence, with competing ideas ignored.

Various authors have argued that persuasion is central to the diffusion of innovation. For instance, Elston (2014) generated two different configurations of the same politico-administrative idea by looking at rhetorical argumentation, narrative voice, and discursive

differentiation. Smith (2013, p. 160) outlined various ways in which ideas can be packaged persuasively, for example, noting that this *"may include employing persuasive metaphors and analogies that make sense to non-experts and cohere with widely held values"* (Chapman, 2007). Similarly, Morrell and Hewison (2013, p. 1) argue that policy texts use rhetorical devices to present policies in the *"best light to justify the ownership of power".* Smullen (2010) considers how the 'New Rhetoric' framework for analyzing argumentation assists with speaking persuasively about public management reforms in different political cultures. Various studies have also focused on a classical rhetoric concept (the 'enthymeme') which focuses on implicit meaning (e.g. Feldman & Idberg, 2002; Knight & Sweeney, 2007). While different strategies of persuasion exist in the literature and no single list is exhaustive, they serve as a useful starting point for considering why some administrative doctrines *"fit the lock of administrative argument in particular times and places, while others fail to do so"* (Hood & Jackson, 1991, p. 28).

Novelty

There is also a school of thought that holds that people and organizations pursue ideas that are perceived as novel. Ultimately, our culture's fascination with change has contributed to the *"ever-present search for new styles, fashions and fads";* the search is ongoing due to the desire to *"replace the dated style of 'yesterday's management' by a new look doctrine, with accompanying special argot"* (Spann, 1981, p. 14). Individuals and organizations *"often find advantages in hooking their wagons to some rising enthusiasm"* (Best, 2006, p. 92). Pursuing novel ideas can signal to the public that the organization is constantly improving, reforming and innovating. Even the process of "*becoming associated with a trendy novelty suggests to others that you are with it, on top of things, in the know, progressive, forward looking"* (Best, 2006). As a result, *"the confidence which is a prerequisite of legitimate*

action is renewed" (Brunsson & Olsen 1993, p. 175). Conversely, those who resist reforming are *"easily labelled old-fashioned, outmoded, obsolete, obstructive, irrational or reactionary"* (Brunsson & Olsen 1993, p. 16*).*

2.3.2 Why organizations do not innovate

In the context of innovation, the public sector has long been perceived as being different from the private sector; thus, there are specific reasons why public sector organizations do not innovate (Bozeman & Kingsley, 1998; Mulgan & Albury, 2003; Bloch & Bugge, 2013; Koch et al., 2006). Broadly, the public sector has often been viewed as conservative, bureaucratic, risk-averse, poor at innovating compared to the market sector and a passive recipient of private-sector innovations (Bloch & Bugge, 2013; Moran, 2010; Potts & Kastelle, 2010). An overarching assumption is that the public sector suffers an innovation deficit or challenge, largely due to weak incentives to innovate (Grady, 1992; Borins, 2001b; Walker, 2003; Albury, 2005; Hartley, 2005; Mulgan, 2007; Windrum & Koch, 2008; Maddock, 2008). While private sector firms are required to take risks to survive in a competitive market, public sector organizations arguably have less to gain from taking risks and less to lose for failing to take risks (Vivona et al., 2020). A failure to innovate within the public sector rarely produces 'devastating' consequences, resulting in innovation being easily discarded when time or budgetary pressures arise (Potts & Kastelle, 2010). Due to possible equity issues in social service delivery, experimental learning is not necessarily encouraged in the public sector (Potts & Kastelle, 2010). The risks associated with unsuccessful innovation are likely more significant for the public sector due to factors such as the media and opposition parties (Borins, 2001a; Potts & Kastelle, 2010).

While a culture exists within the private sector which rewards successful innovators, this is far less common in the public sector (Borins, 2001a). Failed innovations can leave policy issues unsolved or exacerbated, and prevent scarce time and resources from being devoted to more certain, likely incremental, policy solutions (Elston & Bevan, 2020). Decisions resulting in poor outcomes attract excessive blame, whereas decisions resulting in positive outcomes often receive minimal credit from the delegator (e.g. voter) (Elston and Bevan 2020). Avoidance of failure is thus prioritized in the public sector, with risk-taking through experimentation often being institutionally blocked (Potts & Kastelle, 2010; Vivona et al., 2020). Yet it is also possible that public sector innovations occur to counteract the above perceptions/expectations.

2.3.3 Why organizations abandon innovations

It is also essential to consider why organizations abandon ideas, noting that organizations are not *"passive receptors of legitimate ideas"* (Gondo & Amis, 2013, p. 229; Oliver, 1991); and *"a range of conformity, or shades of rogueness exist across organizational fields"* (Quirke, 2013, p. 1692). The adoption and subsequent rejection of an administrative reform can occur for various reasons; for example, an organization may learn that it is inefficient, or *"its faddish appeal"* may decline (Abrahamson, 1991, p. 599; Abrahamson, 1986). Potential adopters of an innovation *"rely on the actions of others to infer the value of a course of action because they are uncertain of its value*" (Rao et al., 2001, p. 521). Conversely, once an organization has already adopted an idea, it can make its own direct evaluations and does not need to use external cues to make abandonment decisions. Thus, abandonments are likely to be linked to an organization's assessment of the results rather than mimicking. Ideas that have been adopted due to mimicking are particularly fragile and are often later judged to be erroneous (Rao et al., 2001). Compared to 'fanfare adoption,

"disillusionment and the abandonment of novelties that aren't working well tend to be much quieter processes; nobody wants to publicize their mistakes" (Best, 2006, p 160). Nonetheless, there may be 'loyalists' who are reluctant to abandon the innovation, holding that it *"could have, should have worked, they insist, but it wasn't given a decent shot"* (Best, 2006, p. 122). Thus, the (known) level of abandonment may not reflect the actual level of satisfaction or disappointment with the innovation.

2.4 Outcomes of innovation

The diffusion of innovation can have a range of outcomes. Rogers (2003, p. 436) emphasizes this when classifying the 'consequences' of innovating as desirable versus undesirable; anticipated versus unanticipated; and direct versus indirect. The management fashions literature recognizes that while some fashionable ideas are 'trivial' and/or largely 'cosmetic', others can *"have massive, sometimes helpful, but sometimes devastating, effects on large numbers of organizations and their employees"* (Abrahamson, 1996, p. 280). Greenhalgh et al., (2004, p. 617) make it clear that both "good" and "bad" ideas spread and pose the following question: *"how can we identify "bad ideas" likely to spread so that we can intervene to prevent this?"* They also ask, *"What are the characteristics of organizations that successfully avoid taking up "bad ideas"? Are they just lucky, or do they have better mechanisms for evaluating the ideas and anticipating the subsequent effects?"* (Greenhalgh et al., 2004, p. 619). In the sections below, I discuss the possible outcomes of innovation.

2.4.1 Social interventions often prove to be disappointing and are poorly measured (if at all)

Policy ideas are often pursued with great enthusiasm and poorly measured, potentially leading to the subsequent adoption of similar programs (Baier et al., 1986; Brunsson & Olsen., 1993; Craig, 2006; Hood & Jackson, 1991; Hood & Dixon, 2015; King & Crewe.,

2014; Keeling, 1972; Taylor & Balloch, 2005). One reason for the persistent over-selling of ideas is the extent of inertia in the public sector. In this context, policymakers often need to give an idea a considerable 'push' to secure even a small amount of support. Examples of other explanations include naivety (i.e. assuming that an idea works) and the vested interests of consultants. Reform ideas are 'packaged together' by their advocates with the elements of *"novelty, explanation, evidence, mystery, breadth, a recipe, and status"* (Best, 2006, p. 58), thus being conveyed as a plausible way of dealing with an intractable policy problem. Even in situations where the government has made strong rhetorical commitments to evidence-based policy (e.g. New Labour), research evidence has rarely been drawn upon (Blunkett, 2000; Cabinet Office, 1999a, 1999b; Smith, 2013a). Impact evaluations (if conducted) are often based on the analysis of rhetoric and ideology, rather than careful measurement of the tangible costs and benefits (Hood & Dixon, 2015). For most government spending, *"we have no idea how effective it is"*, as *"performance is rarely assessed and measurement tends to focus on tracking the number of people served and the amount of service provided, rather than the outcomes that are achieved"* (Liebman, 2013, p. 3). Beliefs about the effectiveness of an innovation can be more important than knowledge of actual outcomes (Dearing & Cox, 2018).

2.4.2 Pursuing an innovation may be beneficial regardless of results

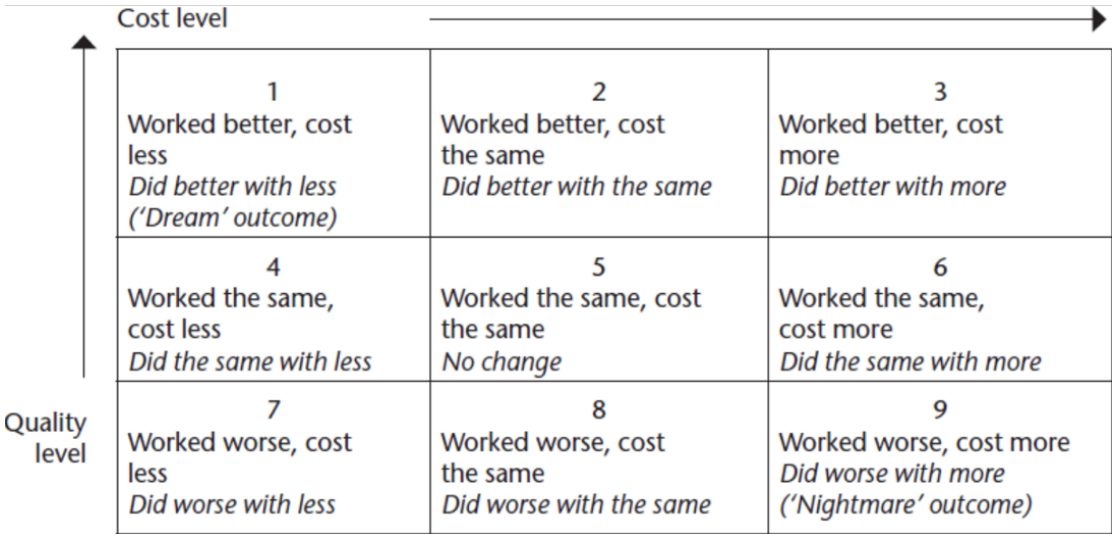
Fads or fashions can fulfil a symbolic function, such as signalling innovativeness, without boosting performance (Abrahamson, 1986; Fombrun, 1987). By projecting an image of innovativeness, organizations can benefit, for instance, by attracting customers or raising capital from other organizations (Nystrom & Starbuck, 1984). Thus, even when a fad or fashion has few political or symbolic efficiencies, it should not be assumed that it is harmful. Administrative fads and fashions can also play an integral role in drawing attention to

problems and solutions that may have otherwise been overlooked. Abrahamson (1991, p. 609) notes that *"organizations may have no choice but to adopt and reject many technically inefficient innovations while searching for a few efficient ones".* The cost of continually adopting and rejecting innovations to find a technically efficient idea, could be much lower than the returns from the efficient innovation. Consequently, fads and fashions may represent critical processes that *"animate random variations from which increasingly efficient innovations can evolve"* (Abrahamson, 1991, p. 609; Anderson & Tushman, 1991; Hannan & Freeman, 1977).

2.4.3 There are high opportunity costs associated with innovation

Allocating scarce resources to a particular innovation incurs opportunity costs. The resources could have been put towards a different policy issue or program; even if the intervention was effective, another program could have been more effective at meeting policy objectives. For instance, in the context of the London Underground Public Private Partnership (PPP), various (cheaper and less complex) alternatives were available (e.g. issuing traditional government-backed bonds rather than resorting to private sector borrowing) (King & Crewe, 2014). Likewise, in the context of SIBs, even if a project succeeds at the program level (e.g., reduces recidivism rates), it is possible that the same outcomes could have been achieved without the complex, expensive funding model. Opportunity cost relates to tangible resources such as funding, along with political goodwill and energy. For instance, Brexit was perceived as an issue that was *"absorbing the oxygen needed to solve other problems"* (Cavendish, 2019). Various permutations of better, worse and neutral outcomes can result from pursuing new ideas. While in some instances, a policy change could 'work the same' and 'cost the same', it is possible that a new idea could work worse and cost more, leading to a 'nightmare outcome' (Figure 2.5) (Hood & Dixon, 2015).

Figure 2.5: Nine Possible Permutations of Better, Worse, and no Change Outcomes



Source: Hood and Dixon (2015)

2.4.4 There is a strong selection bias towards successful ideas

Innovations that have diffused widely or been wholly abandoned are often prioritized in studies, thus introducing a form of selection bias (Strang & Soule, 1998). In considering the outcomes of innovation, it is, therefore, integral to recognize that we often only see the outcomes for innovations that sufficiently "take off" or are complete failures. Due to the clear focus on ideas that have become policy, we do not hear about the path not taken, leading to heavily biased results. This bias is evident in the case of SIBs; for instance, various tweets focused on SIBs have mentioned projects that never launched, such as a UK drug repurposing program for rare diseases and a SIB to assist with funding Australia's National Disability Insurance Scheme (Cashin-Garbutt, 2016; Conifer & McKinnon, 2016). There are many completed SIB projects for which no evaluation is available.

2.4.5 Pro innovation bias emerged in the early literature

The early innovation-diffusion literature focused on three overarching questions, as identified by Abrahamson (1991, p. 586), Rogers, 1962, 1983; Rogers & Schoemaker, 1971). Firstly, *"what processes and contextual factors affect innovations' rates of diffusion?"* This question focused on identifying whether mathematical models accurately describe long-term changes in diffusion rates (Rogers, 1983). Secondly, *"what characteristics differentiate earlier from later adopters"?* This strand of research categorized adopters according to the timing of their uptake, to identify any differences between 'leaders' and 'laggards' (Ryan & Gross, 1943; Rogers, 1983). Finally, *"how does the structure of networks of adopters affect the sequence in which adoptions occur during diffusions"?* This question was answered using network analysis (Rogers, 1983).

These questions received extensive attention in the early literature, as *"answering them appeared highly practical during the post World War II era of US economic dominance"* (Abrahamson, 1991, p. 587). As high rates of innovation fuelled economic growth during this time, pro-innovation biases developed in the diffusion of innovation literature, defined as *"presumptions that innovations will benefit organizations"* (Kimberley, 1981; Abrahamson, 1991, p. 589). This resulted in the viewpoint that innovations diffuse when they benefit the organizations adopting them and disappear when they are not beneficial (Downs & Mohr, 1976; Van de Ven, 1986; Zaltman et al., 1973). Subsequent research thus focused on helping to spread seemingly beneficial innovations, covering topics such as how to speed up diffusion rates; how to overcome the resistance of laggards' to adopting innovations; and how to intervene in diffusion networks (Abrahamson, 1991).

2.4.6 Researchers have now introduced an anti-innovation bias

By attempting to overcome the existing pro-innovation biases in the literature, researchers introduced anti-innovation biases. While the early literature contained pro-innovation biases, a more sceptical view of innovation subsequently emerged (Kimberly, 1981). This view included claims that management "fads" or "fashions" have continuously diffused technically inefficient innovations, thus causing technically efficient innovations to be rejected. Thus, it is now important for researchers to understand how diffusion processes can lead to organizational benefits, in addition to possible harm. On the one hand, 'anti-innovation' biases could result in institutions viewing most innovations as fads and thus not adopting them, even if they are efficient. Conversely, fads or fashions could prompt the diffusion of inefficient innovations, along with the rejection of efficient innovations. Recognizing this, Abrahamson (1991, p. 587) notes that researchers should spend less time on questions such as "*What affects diffusion rates?"*; and more time on questions such as *"When and by what process are technically inefficient innovations diffused or efficient innovations rejected"*. Focusing on the latter would provide more useful information to decision-makers who benefit from knowing whether or not the most technically efficient innovations are being diffused and retained, in addition to understanding diffusion rates.

2.5 Conclusion

In this chapter, I have outlined what is known about the diffusion of early-stage public sector innovations. I centred this chapter around three sections, including: 1) definition, types, and stages of innovation; 2) drivers of innovation; and 3) outcomes of innovation. In each of these sections, I discussed the relevance of the concept of uncertainty, demonstrating that it matters in the context of public sector innovation and is thus a persistent and vital theme in this thesis. In relation to the innovation stages, I noted that the initial stages (knowledge,

persuasion, decision) are the most relevant for this thesis as I am aiming to better understand why some early-stage innovations, for which there is the greatest level of uncertainty, rapidly spread throughout the public sector. Regarding the drivers of innovation, I provided various reasons as to why organizations innovate, ranging from rational explanations (e.g. seeking a solution to address a problem) to less rational drivers such as novelty, persuasiveness and the powerful role of management consultants. Finally, I also discussed why public sector organizations do not innovate (e.g. inertia, risk averse); and why organizations may adopt but then abandon ideas.

In the next chapter, I present the theoretical framework for this thesis. I argue that SIBs provide an example of organizations pursuing an innovation for reasons other than efficiency benefits (particularly legitimacy reasons). Thus, it is integral to consider how factors such as norms, belief systems and cultural pressures have contributed to the spread of SIBs. Institutional theory provides the most useful lens for analyzing my overarching research question, as it accounts for the influence of these factors, rather than only instrumental factors such as efficiency gains. Notably, the theme of uncertainty, which has been explored throughout this chapter, will also be woven into the theoretical framework, and discussed in the empirical chapters of this thesis.

Chapter 3: Theoretical Framework

3.1 Overview

In this chapter, I outline the theoretical framework for analyzing why early-stage, unproven innovations diffuse and what truncates this, leading to "thin" rather than "thick" institutionalization. In this thesis, I primarily use the lens of organizational institutionalism to derive predictions about how and why SIBs have spread to various public sectors. The orthodox, instrumental view holds that organizations sharing the same environment take on similar forms as "efficiency-seeking entities who are seeking the optimal fit with their environment*"* (Brunsson & Olsen, 1993; Etzioni, 1964). I argue, however, that this view is a naive and inadequate way of explaining the diffusion of SIBs; a powerful complement and critique is institutional theory. As with many existing organizational studies, I compare the instrumental (orthodox) view to the institutional view and consider how a nuanced combination of these account for the spread of SIBs over time. I argue that the interplay of efficiency and legitimacy drivers has changed over time as levels of uncertainty about the success of the SIB innovation have changed. While instrumental motives may have driven initial SIB projects, I argue that legitimacy motives have become increasingly important as negative evidence has accumulated in the literature. However, legitimacy can only carry an idea so far; the diffusion of SIBs has thus been truncated as the novelty has worn off, and variations such as DIBs have been introduced to sustain interest. The sections below explain the theoretical framework in more detail.

3.2 Central Hypothesis

For this DPhil thesis, diffusion (i.e. SIBs adoption) will be the dependent variable, while legitimacy will be the key independent variable. For conceptual clarity, I will define the key terms of 'diffusion' and 'legitimacy' before expanding on my hypothesis. As I am

interested in the link between evidence of performance, legitimacy and the diffusion of innovations, I will also define relevant terms such as 'effectiveness', 'evidence of performance' and 'efficiency'. Diffusion is *"the spread of something within a social system"*, from a source to one or more subsequent adopters (Strang & Soule, 1998, p. 266). This may include the spread of structures, behaviours, beliefs and/or technologies (Frederickson et al., 2016). Legitimacy is *"a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions"* (Suchman, 1995, p. 574). This definition recognizes that legitimacy is an umbrella perception or assumption that is created subjectively. It is socially constructed *"in that it reflects a congruence between the behaviours of the legitimated entity and the shared (or assumedly shared) beliefs of some social group"* (Greenwood et al., 2017, p. 359). Thus, it depends on a collective audience and represents a reaction of observers to the organization, as they perceive it (Greenwood et al., 2017, p. 359).

As with legitimacy, there are various definitions for the term efficiency and effectiveness. Effectiveness typically refers to the ability of an intervention to have a meaningful impact on recipients, under 'normal' (rather than controlled) conditions, while efficiency refers to the relation between resources and outcomes, with the goal being to make the most economical decisions (Burches & Burches, 2020). Many public administration scholars insist on maintaining a clear distinction between these two concepts, as *"effectiveness refers mainly to goal-fulfilment, and efficiency to output per unit of input"* (Deva, 1985: 94). Given the data limitations that are inherent in this field, it can be challenging to demonstrate the effectiveness of SIBs, let alone efficiency. Nonetheless, for this thesis, the terms efficiency and effectiveness will be kept separate, although both are relevant when considering whether or not an innovation is technically functional in terms of

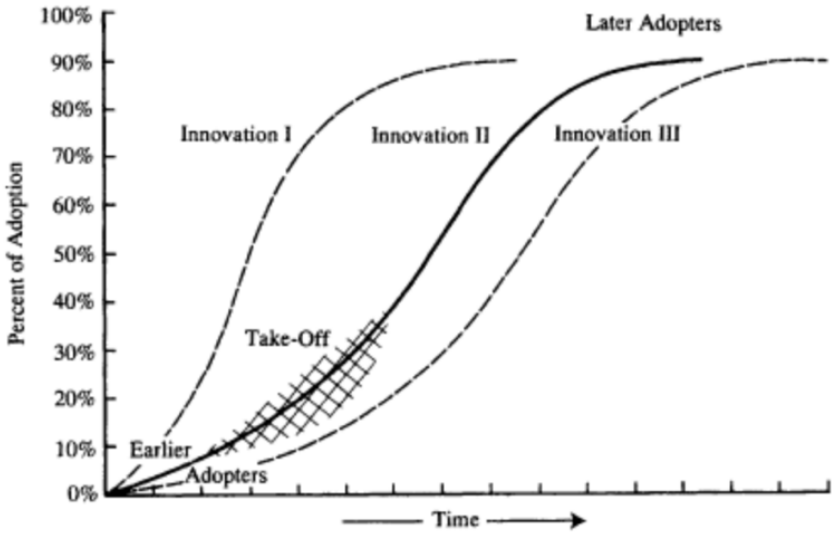
living up to claims about what it should do (Schuck 2014). For example, noting that SIBs typically have high transaction costs, even if a SIB-funded program is effective, it may not be efficient as it is always possible that the same outcomes could have been achieved far more efficiently, without the SIB funding model.

Assessments of program effectiveness (or otherwise) depend on how an individual defines and measures failure and success. These judgements are contestable, particularly given that some programs are *"successful in some respects but not others"*; and *"even ineffective ones create some benefits for at least some people*", who may therefore consider it to be successful (Schuck 2014, p. 6). While supporters of a program often perceive great success, opponents often perceive failure. Noting this subjectivity, Schuck (2014, pp. 4-5) emphasised that *"the main index of a program's performance should not be its durability or its enthusiastic defenders, which may reflect political inertia protected by strategically positioned beneficiaries"*. Rather than holding public policies to 'impossible standards', it is important to use well-recognised criteria *"which in principal are hard to dispute"*, such as quantitative systematic review methods, which I draw upon in Chapter 5 of this thesis (Schuck, 2014, p. 41). When referring to the term evidence (and/or evidence of performance) in this thesis, I will be referring to actual empirical evidence rather than only perceptions of performance. Although perceptions matter, it is well understood that particular types of evidence are more robust than others (Parkhurst & Sudeepa, 2016).

In this thesis, levels of legitimacy are predicted to affect different rates of diffusion and/or institutionalization. I argue that legitimacy matters in the diffusion of public sector innovation; ideas carrying legitimacy benefits are likely to be adopted relatively rapidly, as with Innovation I (Figure 3.1). While the existing literature notes that the perceived attributes

of an innovation (e.g., relative advantage or compatibility) can affect the rate of adoption (Rogers, 1995), further empirical research is required to better understand how legitimacy affects adoption rates in the context of early-stage public sector innovation. Of particular interest, is what happens to legitimacy and/or efficiency motivations when the level of uncertainty regarding intervention effectiveness changes over time. While legitimacy is expected to be the main driver, it will also be important to explore and discount other possible explanations. Thus, I will empirically test the extent to which instrumental or institutional (or a combination of both) drivers have influenced the uptake of SIBs.

Figure 3.1: Different Rates of Adoption



Source: Rogers (1995)

The central hypothesis described above is integral as current theories of diffusion do not necessarily account for situations in which the cumulative number of adoptions is still too low to trigger isomorphic responses, and high uncertainty and/or negative experiences of early adopters would discourage further adoption, when based on rational assessments

(Compagni et al., 2015). Institutional theories suggest that innovations diffuse initially due to their technical or economic benefits, but, once a critical mass has been reached, isomorphic pressures are triggered (Abrahamson, 1991; Tolbert & Zucker, 1983). However, in the SIBs context, these assumptions are problematic as they do not account for the fact that the relative performance remains uncertain more than 12 years on from the first project launch and the innovation has not spread as widely as first anticipated, suggesting that the number of adopters may not have reached the threshold required to trigger isomorphic pressures (Compagni et al., 2015; Roy et al., 2017).

3.3 The instrumental and institutional views

In exploring what has underpinned the widespread growth of SIBs, I will contrast the instrumental (efficiency-driven) view with the institutional (legitimacy-driven) view. These two viewpoints have been applied in the existing literature on the diffusion of innovations to understand adopter motivations (e.g. Tolbert & Zucker, 1983). The instrumental view builds on the rational actor model and assumes that the decision to adopt an innovation is motivated by a desire for technical benefits, such as increased efficiency (Katz & Shapiro, 1987; Teece, 1980). The institutional view represents a sociological viewpoint and holds that adoption stems primarily from the desire to appear legitimate (Abrahamson, 1991; DiMaggio & Powell, 1983). It recognizes that organizations not only respond to technical demands and pressures but also to what they believe society expects of them. These two perspectives are discussed below.

3.3.1 Instrumental view

When considering what explains the rapid rise of particular innovations, such as SIBs, it is important to firstly consider what is already known about the instrumental perspective.

Instrumental thinking is the dominant, orthodox way of theorizing about organizations and performance. Studies of organizations need not use the label 'instrumental' to be considered instrumental. Unless the theory is critical or explicitly institutional, it can be regarded as instrumental as it explains organizational improvement through rational decision-making. Desired outcomes and beliefs about cause-and-effect relationships are central to the instrumental perspective (Tsoukas & Knudsen, 2005). Organizations are viewed as rational actors (although in complex environments) who adopt institutional forms and/or procedures for efficiency reasons (Thompson, 1967; Blau & Shoenherr, 1971). It is assumed that organizations are consciously searching for the most efficient means for reaching organizational goals (Thompson, 1967; Blau & Schoenherr, 1971; Scott & Davis, 2007). Organizations are seen as *"the product of an intensive effort to devise ever-more efficient structures for performing the tasks associated with modern society"* (Hall & Taylor, 1996; Scott, 1987). Ultimately, a 'chain of cause and effect' runs from policymakers' intentions, to new structures and processes, to changes in behaviour and improved results. The instrumental perspective, therefore, points to the need for effectiveness or efficiency following adoption (Tolbert & Zucker, 1983).

Early instrumental research emphasized that the essence of this approach is *"whether the specified actions do in fact produce the desired outcome, and the instrumentally perfect technology is one which inevitably achieves such results"* (Thompson, 1967, p. 14). Thompson (1967, p. 159) noted that 'rationality' and 'uncertainty' are binary opposites, with uncertainty appearing "*as the fundamental problem for complex organizations"*. Uncertainty is perceived as representing *"darkness and undesirability"*, with the amount of energy devoted towards moving to certainty linked to how tight (or otherwise) the norms of rationality are (Tsoukas & Knudsen, 2005, p. 201). Linked to this, Scott (1998, p. 33) noted

that organizations are instruments specifically designed to attain specified goals; 'rationality' in this context is used narrowly to refer to the *"extent to which a series of actions is organized in such a way as to lead to predetermined goals with maximum efficiency".* While organizational scholars have primarily focused on the search for *"techniques that link ends and means in the most efficient way"* (Tsoukas & Knudsen, 2005, p. 200), there are problems inherent in elevating instrumental rationality to a 'supreme position' without *"introducing political, cultural, and moral balances"* (Tsoukas & Knudsen, 2005, p. 200; Weber, 1947, 1949). Thus, an instrumental view should not necessarily be the *"sole arbiter of reality"* (Tsoukas & Knudsen, 2005, p. 200).

Instrumental innovation

Compared to institutional innovation, instrumental innovation assumes that innovation proceeds along a strategic planning process. According to the instrumental view, successful innovation is *"the result of (a) careful planning of a superior product for an attractive market and (b) the execution of that plan by a competent and well-coordinated cross-functional team that operates with (c) the blessings of senior management"* (Brown & Eisenhardt, p. 348). This involves several strategically planned steps such as information gathering, analysis, evaluation and action (Cooper, 1993). The steps can range from initial screening to feasibility studies, trial production and market launch. Thus, the rational process is highly complex, particularly when considering the dynamic environments that innovations are launched in. Regardless of the complexity, rational models of innovation are evident in many organizations and research projects. In summarizing evidence on successful and failed innovations that have followed a rational model, Brown and Eisenhardt (1995) listed the following as the main factors for successful, rational product innovation: 1) product advantages; 2) market attractiveness; and 3) internal organization. While product innovation

differs from public sector innovation, these factors assist with framing instrumental innovation around the technical benefits gained by pursuing an innovation.

Links to the evidence-based movement

As the instrumental view centres around desired outcomes and beliefs about cause-and-effect relationships, it is important to consider how it links to the evidence-based movement. The Evidence-Based Medicine (EBM) movement has gained increased prominence since the early 1990s as an approach to clinical care decision-making by physicians (Greenhalgh, 2014; Straus et al., 2018). It involves *"the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients"* (Sackett et al., 1996, p. 71). Within this, there are two particularly essential methodological principles (Oakley et al., 2005). Firstly unbiased, controlled comparisons are prioritized in evaluating the effectiveness of interventions. Secondly, there is a strong need for timely, comprehensive syntheses of existing effectiveness evidence. The rise of EBM has led to the promotion of "Evidence-Based Practice" (EBP) in other areas, including social services (Nutley et al., 2000).

In the UK, evidence-based policy making (EBPM) gained prominence in the context of the Labour Government's commitment to modernize government (Nutley et al., 2000). There was a deliberate push to ensure that public services became more responsive to users' needs and were more efficient, effective and delivered to higher quality standards (Cabinet Office, 1999a). Cabinet Office (1999b) emphasized that a critical driver of the proposed modernization was evidence-based policy making and service delivery, as 'what matters is what works'. EBPM in this context means that *"policy initiatives are to be supported by research evidence and that policies introduced on a trial basis are to be evaluated in as*

rigorous a way as possible" (Plewis, 2000, p. 96). The central assumption of EBPM is that policy action taken by the government is rational to the extent that it is based upon 'sound evidence'. As EBPM traces back to the EBM movement, it has inherited various features of EBM, which revolve around outcomes measurement. This includes the prominence of systematic approaches, including systematic review methods. Consequently, several authors have explored instrumental rationality's role in EBPM. For instance, Sanderson (2002, p. 62) analyzed the extent to which the increased enthusiasm for the use of evidence in policymaking *"is indicative of 'technocratic politics' underpinned by an instrumental rationality which erodes the normative basis of policy making and undermines the capacity for 'appropriate' practice".*

Just as scholars have highlighted the problems inherent in solely using an instrumental viewpoint, several academics have criticized EBPM (Gibbs & Gambrill, 2002; Greenhalgh & Russell, 2009). In particular, there is concern that it removes the human element from clinical practice. Hammersley (2005) noted that EBP can result in overlooked contextual factors and that practice should be a matter of judgement, not just research. Similarly, several critics have noted that evidence-based approaches ignore the complexity of the policymaking process, presupposing a 'linear' process that has long been perceived as unrealistic (Black, 2001; Sabatier & Weible, 2014). Webb (2001, p. 67) suggested that it is a 'cookbook' approach *that "ignores the processes of deliberation and choice involved in . . . decision making."* Various criticisms have also centred around the evidence 'paradox' that exists, with proponents calling for high-quality empirical research yet often failing to meet their own standards. Specifically, Hammersley (2005) noted that research is fallible as systematic reviews rely on randomized controlled trials (RCTs), yet RCTs are often poorly implemented. Similarly, Webb (2001, p. 62) critiqued the fact that the only 'admissible' form of evidence in this field

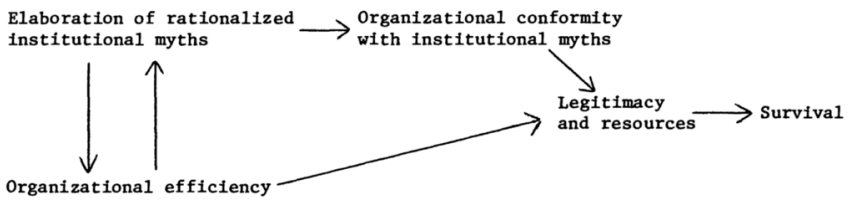
is an RCT. Despite criticisms such as these, the concept of EBPM continues to attract much policy and academic attention (Smith, 2013a; Smith, 2013b). This attention reflects the continued dominance of the instrumental view.

3.3.2 Institutional view

While the starting point for considering why an innovation has diffused is often the instrumental perspective, the institutional perspective provides an important and necessary counterweight. Institutionalists hold that legitimation is central to the spread of ideas (Hall & Taylor, 1996; Meyer & Rowan, 1977). Thus, a central concept within organizational institutionalism is legitimacy theory. Legitimacy influences the behaviour of organizations and affects performance and survival (Pollock & Rindova, 2003; Singh et al., 1986; Greenwood et al., 2017). It can be valuable for various reasons; for example, it assists with resource acquisition and reduces external oversight as it *"protects the organization from having its conduct questioned"* (Meyer & Rowan, 1977, pp. 349-350). Issues of perception are central to the concept of legitimacy. Consequently, there can be multiple co-existing and competing notions of what is acceptable and 'legitimate' (Greenwood et al., 2017). Whose legitimacy is considered to be valuable can be reflective of power dynamics. Organizational institutionalists do not necessarily view individuals as being irrational or as failing to be goal-oriented. Instead, they emphasize that what an individual perceives to be a rational decision is socially constructed (Greenwood et al., 2008).

There are various sources of legitimacy, including performance (i.e. output) legitimacy or behaving as expected by external constituencies (Figure 3.2). In policy areas where output legitimacy is difficult to demonstrate (e.g. SIBs), governments may be more likely to rely on the latter (i.e. behaving as expected).

Figure 3.2: Sources of Legitimacy – Organizational Survival



Source: Meyer and Rowan (1977)

Legitimacy has different characteristics and dynamics that inevitably change over time (Binz et al., 2016). Regardless, the concept of legitimacy has remained core to institutional theory across time and has driven many studies focused on examining the relationship between organizations and their institutional context (Scott, 2013; Suchman, 1995). From these works, four pillars of legitimacy have emerged, including: 1) normative; 2) cognitive; 3) regulative; and 4) pragmatic (Table 3.1). Each of these reflects a different aspect of legitimacy, providing a useful framework that can be applied when studying legitimation in empirical studies (Binz et al., 2016; Scott, 1995).

Table 3.1: Key Dimensions of Legitimacy

	Cognitive	Normative	Pragmatic	Regulative
Motivation	Taken-for-grantedness	Social obligation	Self-interest	Expedience
Affect	Certainty/confusion	Shame/honour	Utility/indifference	Fear, guilt/innocence
Basis of legitimacy	Comprehensible, culturally supported	Morally governed	Personal evaluation	Legally sanctioned

Source: Binz et al., (2016, p. 251) - adapted from Scott (2008) and Suchman (1995)

The various pillars do not form a hierarchy; the elements can change over time, and within a specific timeframe or context and often overlap or co-exist (Scott, 1995; Suchman, 1995). Cognitive legitimacy refers to the spread of knowledge about a new innovation among

social audiences (Aldrich & Fiol, 1994). There are two types of cognitive legitimacy, namely: (1) legitimacy based on comprehensibility; and (2) legitimacy based on taken-for-grantedness (Suchman, 1995). The former relates to understanding the 'working principles' of an innovation (Scott, 2013; Jansma et al., 2020, p. 197), while the latter occurs when the innovation becomes a crucial, taken for granted part of the relevant institutional framework (Kaganer et al., 2010). Normative legitimacy involves judgements about whether an innovation is appropriate for society. This assessment is based on the promotion of social welfare, taking into account values and norms. Pragmatic legitimacy revolves around the self-interested calculations of the most immediate audiences of an organization (Suchman, 1995; Kaganer et al., 2010; Walker et al., 2014). It involves the evaluation of the innovation's utility as it focuses on tangible benefits (Golant & Sillince, 2007). Regulative legitimacy refers to the alignment of new practices with the relevant laws, rules and regulations within a particular field (Kaganer et al., 2010; Scott, 2001; Zimmerman & Zeitz, 2002). Innovations that have such alignment possess a higher level of regulative legitimacy, than those that do not. The various pillars are likely relevant to the context of SIBs to differing degrees. They will primarily be applied in the qualitative systematic review (Chapter 6), although they will also be briefly applied and discussed in relation to the (random sample) dataset contained in Chapter 8.

Institutional innovation

As this thesis focuses on public sector innovation, it is important to consider, more specifically, what is meant by 'institutional innovation'. From an institutional view, innovation is *"a tactic to gain legitimacy with [these] organizations, rather than, necessarily, an end in itself"* (Osborne, 1998, p. 52). Noting that innovation has *"diffused in all areas of life and mutated to a highly desirable good",* it can assist with legitimacy by ensuring that an

organization looks attractive to funders and broader stakeholders (Institute for Nonprofit Management, n.d). While innovation may refer to a product or process, it is *"sometimes only rhetoric, and an attitude which is […] manifested in ideals of modern Western societies"* (Brandl & Bullinger, 2009). The fact that the attribute of being innovative is often preferred to being efficient or effective is highlighted by the emergence of various innovation awards and innovation focused grants or support programs (e.g. EQUAL) (Institute for Nonprofit Management, n.d).

Developments in the institutional literature

Early institutional works explored how and why organizations often resemble each other, over time, due to the strength of legitimacy and the various socio-cultural pressures (Meyer & Rowan, 1977). Within the early literature, there was a well-recognized efficiency-legitimacy trade-off, noting that *"conformity to institutional rules often conflicts sharply with efficiency criteria and, conversely, to coordinate and control activity in order to promote efficiency … sacrifices its support and legitimacy"* (Meyer & Rowan, 1977, p. 341). More recent studies have held that efficiency and legitimacy are not necessarily mutually exclusive. Regardless, it is likely that a trade-off still exists to some extent in many situations. Figure 3.3 depicts the associated risks; for instance, decision-makers may overlook efficient ideas that are illegitimate and pursue inefficient ideas that are legitimate (orange boxes).

Figure 3.3: Possible Risks Introduced by Efficiency-Legitimacy Trade-off

		Efficient?	
		Yes	*No*
Legitimate?	*Yes*	**Efficient** & **legitimate**	**Legitimate** but **inefficient**
	No	**Efficient** but **illegitimate**	**Inefficient** & **illegitimate**

Source: Elston, forthcoming, Understanding and Improving Public Management Reforms, Bristol: Policy Press

The notion that evidence is de-prioritized in adoption decisions and/or evidence is itself socially constructed is strongly aligned with the emergence of new institutionalism in the 1970s. At this time, scholars began to accept the fact that organizational forms and decisions have *"little to do with the objective tasks in which organizations were engaged and much to do with the social relationships in which they were embedded"* (Greenwood et al., 2008, p. 758). This was a significant shift as organizational studies had been dominated, in the post-WWII economy, by a managerialist engineering approach. New institutionalism and its focus on organizational fields has increasingly been critiqued, as it has been argued that *"an adequate theorization of institutional dynamics—including the structuration of fields—required an interinstitutional conceptualization of society"* (Friedland & Alford, 1991; Lounsbury et al., 2021, p. 262). Friedland and Alford (1991, p. 248) emphasized that Western society comprises different institutional orders (e.g. bureaucratic state, democracy, capitalist market etc); and that each of these *"has a central logic—a set of material practices and symbolic constructions—which constitutes its organizing principles and which is available to organizations and individuals to elaborate".*

Building on this, Shipilov et al., (2010) developed a theory of multi-wave diffusion which helps to explain the diffusion of practices related to a particular logic, over time. Their model, which was tested by analyzing the diffusion of board reform practices in Canada (1999 – 2005), showed that an organization's adoption of practices from a particular institutional logic implies its acceptance of that logic. Ultimately, *"this makes organizational adoption of subsequent practices sharing the same logic more likely, irrespective of these practices' adoptions by the organization's network contacts"* (Shipilov et al., 2010, p. 846).

Shipilov et al., (2010) found that while Tolbert and Zucker's (1983) model may explain the diffusion of a single practice (or a group of practices which diffuse simultaneously), it does not explain the drivers of diffusion for multiple waves of practices that are underpinned by the same institutional logic. Their study highlights the importance of drawing upon the institutional logics literature, when considering why certain public sector innovations spread.

The institutional logics research holds that efficiency and legitimacy are variables that are contingent on which institutional logic is used to interpret the situation. Legitimacy is considered a categorical element on the y axis which varies based on the x axis (e.g. family, market, state, community logics). Thus, what is considered legitimate when applying a market logic may be different from what is considered legitimate, from the lens of a family logic. Institutional scholars have warned against combining categories from different logics (e.g. market and family logics) (Thornton et al., 2012; Friedland & Alford 1991). Shipilov et al's (2010) study is particularly relevant to the SIB context, as multiple 'waves' of similar concepts have emerged, both before and after SIBs. For instance, the Big Society Agenda (which SIBs are part of) and the *"desire of the Coalition to usher in an age of austerity"* benefited from a decade of NPM reforms put in place by the outgoing New Labour Government (Guter-Sandu, 2018, p. 41). Further, the DIBs concept benefited from the logics that were already in place from SIBs. Thus, rather than imitating the behaviour of others, SIB organizations may have been further committing to an underlying set of logics that they had already invested in (Shipilov et al., 2010).

3.4 Applying the two alternate views to the context of SIBs

Having explained the instrumental and institutional views in detail, I will now discuss how each view is relevant to the context of SIBs. I will focus on outlining what we would

expect to see if they partly or wholly explained the diffusion of SIBs.

Instrumental view

If the rapid but thin diffusion of SIBs was caused by organizations acting instrumentally, we would expect decisions regarding the adoption of SIBs to be linked to the emerging evidence base. While there would likely be some uncertainty at the outset (as this is inherent in the context of early-stage innovation), we would then expect results to be robustly measured and uncertainty regarding the effectiveness of the idea to decrease over time. When an idea can be properly observed and measured, diffusion may not imply institutionalization given that the results may lead to disappointment and abandonment (Rao et al., 2001). Thus, applying the instrumental view, if certainty increases regarding intervention effectiveness (i.e. positive results), we would expect to observe subsequent adoptions, in which actors are behaving rationally and pursuing the innovation as they could reasonably expect it to assist with achieving their organizational objectives. If, however, certainty increases about innovation ineffectiveness (i.e. mixed, null and/or negative results), we would expect it to be more 'rational' for actors to decide against adopting the idea; and/or to abandon the idea altogether (if they have already launched a SIB).

The instrumental view also suggests it would be unlikely for uncertainty regarding the effectiveness of the idea to remain for a prolonged period, as this view prioritizes the importance of scientific evaluation and robustly measuring the results. In trying to understand the link between evidence, legitimacy and diffusion, it will thus be important to consider the available data on SIB uptake and abandonments.

Institutional view

Conversely, if the diffusion of SIBs could be explained by institutional theory, we would expect continued adoption even if the emerging evidence is neutral or negative. We would also expect to observe a lower standard of evidence, potentially contributing to prolonged uncertainty, as actors would primarily be pursuing the funding model for reasons other than efficiency benefits and thus may not necessarily prioritize robust measurement. Noting that legitimacy is more likely to drive decisions (e.g. to mimic other organizations) in ambiguous situations, we could reasonably expect legitimacy motives to be present if uncertainty remains for a prolonged time after the initial launch of the SIB concept. While a level of uncertainty is inherent in early-stage public sector innovation, this is more likely to remain when potential social gains are prioritized above technical gains in decision-making. Overall, if institutional motives have driven the adoption of SIBs, we would expect ongoing uncertainty, which would prolong the life of the innovation and prevent it from being completely abandoned.

Institutional motives could also be identified by considering the rhetoric that has surrounded the rise of SIBs and the extent to which different strategies of persuasion have been used by stakeholders. This is not an unequivocal signal, as it is possible to be persuaded about a 'rational', instrumental idea as well as a legitimate idea; also, an idea could be both efficient and legitimate. Regardless, while rhetoric and strategies of persuasion matter to both the instrumental view and institutional view, I argue that there is a dynamic element that needs to be considered. If SIBs worked as intended, the results would be convincing and drawn upon by SIB proponents, thus leading to less of a reliance on rhetoric, over time. Conversely, if persuasive techniques (discussed in section 2.3.1), rather than concrete results, continue to be relied upon, I argue that this could be reflective of institutional, rather than

instrumental, drivers. In particular, the continued use of time-sensitive techniques such as framing the idea of new, many years after the emergence of the idea, could point to institutional drivers. Although rhetoric may allow an idea to outlive the time-period in which the instrumental rationality would say it should demise, rhetoric has its own limits. Thus, I argue that the eventual decline in SIBs usage could be linked to the novelty wearing off and thus, rhetoric becoming less persuasive over time.

3.5 Conclusion

This chapter developed the theoretical framework for understanding the rapid and widespread growth of SIBs. It noted that I use an organizational institutionalist perspective to understand the diffusion of SIBs. I argue that the orthodox, instrumental view does not sufficiently explain the spread of early-stage public sector innovation. An important complement to and critique of this view, is the institutional view, which recognizes the importance of legitimacy considerations. This approach is aligned with calls in the institutional literature for more flexible models of diffusion that account for the complex interplay between efficiency and legitimacy motives.

Prior to this chapter, I had outlined what is known about the diffusion of early-stage public sector innovations. I presented SIBs as an example of an idea that has spread throughout various public sectors despite limited evidence of success. I noted that analyzing this paradox could shed light on what explains the rapid growth of some unproven early-stage public sector ideas. A crucial learning from the literature review was that uncertainty is a persistent and important theme to consider; it influences various aspects of innovation, particularly the drivers and outcomes of innovation. As uncertainty is greatest in the initial stages of diffusion, and then changes, it is integral to consider how this influences the spread

of the idea and the drivers of adoption, over time. I noted that we have a limited understanding of what explains the spread of thinly diffused innovations, particularly those which have had a truncated diffusion trajectory.

Having prepared my research agenda in Chapters 1-3, I will now explain how I will prosecute it. The implications of my literature review and theoretical framework are my methodological plan, which will be discussed in Chapter 4. I will use a multi-method research design, to better understand what has driven the spread of SIBs.

Chapter 4: Research Methodology

4.1 Introduction

This chapter explains the mixed-methods methodology used to understand what explains the rapid and widespread growth of SIBs. This includes: 1) an effectiveness systematic review to understand the changing state of academic knowledge over time and the extent to which this supports the instrumental view; 2) a qualitative systematic review to identify broader, legitimacy-based motivations for SIBs adoption; 3) sentiment analysis and network analysis of social media data to understand how practitioner understanding of the reform changed between 2010-2020; and 4) framework analysis of social media data to identify how rhetorical and discursive techniques have been applied, over time, to promote the uptake of SIBs. This multi-part design rests on a 'consilience' approach, which involves collating evidence from multiple sources, *"no one of which is perfect on its own, but which, if combined, provide more powerful evidence than any one element would do on its own if those various sources point in the same direction"* (Hood & Dixon, 2015, p. 14). Importantly, I incorporated previously overlooked data sources such as Tweets. While the early chapters focused on explaining and justifying my research methods, this chapter outlines the methods used to achieve my overarching objective of understanding what explains the diffusion of SIBs. I broadly present the positive case for each method and provide a brief overview, discussing the specific details and limitations in each empirical chapter.

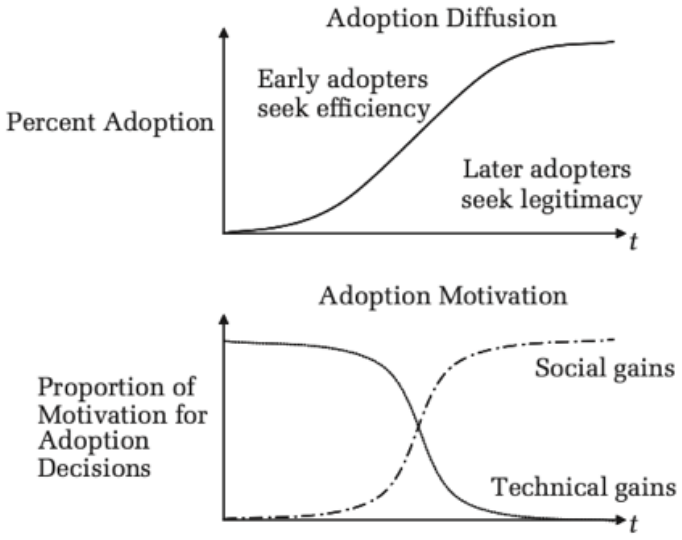
4.2 Existing approaches and their shortcomings

4.2.1 Two-stage model of diffusion

A well-known approach for testing the extent to which adopters seek technical and/or social gains over time is Tolbert and Zucker's (1983) two-stage model of diffusion. As

depicted in Figure 4.1, it holds that early adopters seek technical gains from adoption, while later adopters primarily seek the social benefits that stem from appearing legitimate. In studying the adoption of civil service reform by cities, Tolbert and Zucker (1983) found that city characteristics, such as size, socioeconomic structure and the influx of immigrants, could explain initial adoption. Thus, early adoption was deemed a rational response to the issues encountered by the relevant city (Tolbert & Zucker, 2013; Van de Ven et al., 2013). Further adoption was not linked to instrumental motives; as the innovation had gained social legitimacy, it was then adopted regardless of whether it could offer performance improvements.

Figure 4.1: Adoption Motivations Over Time – Classic Two-Stage Model



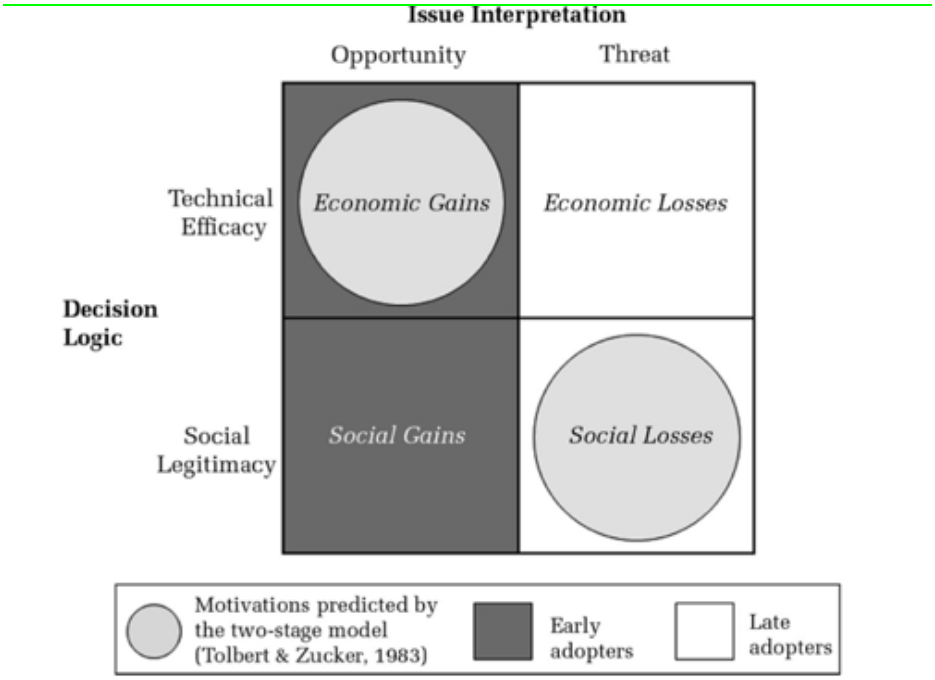
Source: Tolbert and Zucker (1983)

Despite its initial popularity, recent debates have highlighted several concerns with the two-stage model. These include (a) oversimplification – as it oversimplifies the relationship between adoption motivation and timing (Greenwood et al., 2008; Kennedy & Fiss, 2009); (b) overlapping motivations – as social and economic motivations are not

necessarily mutually exclusive (Kennedy & Fiss, 2009; Lounsbury, 2002; Thornton, 2004; Schneiberg & Soule, 2005; Scott et al., 2000); and (c) reliance on theory over data – as many past studies have relied on inference about motivations rather than more direct assessments (Kennedy & Fiss, 2009; Donaldson, 1995; Scott, 1995). The limitations are particularly prevalent in situations where the diffusing innovation has only been partially implemented (Edelman, 1992; Fiss & Zajac, 2004, 2006; Westphal & Zajac, 2001). In the SIBs context, there are additional limitations which cannot be easily overcome, thus necessitating a different approach. Most notably, there is a shortage of predictors of "rational" adoption for this context, as the reform has been adopted in very different contexts, for different policy areas, and the evidence base is still developing.

In response to the well-documented shortcomings, Kennedy and Fiss (2009) found that both early and later adopters are affected by logics of efficiency and legitimacy; and that the importance of social considerations may differ less among early and later adopters than previously hypothesized. Specifically, as depicted in Figure 4.2, they found that early adoption is associated with opportunity framing and motivations to achieve economic and social gains; conversely, later adoption is associated with threat framing and motivations to avoid economic and social losses.

Figure 4.2: Example of an Alternate Diffusion Model



Source: Kennedy and Fiss (2009)

While this model is helpful in highlighting the shift towards more flexible models of diffusion, it contains well-documented pro-innovation biases, as it implies that early adoption involves economic and social gains, while late adoption involves economic and social losses. It also does not account for the impact of changing levels of uncertainty, regarding intervention effectiveness, over time. In analyzing what has underpinned the growth of SIBs, this thesis builds upon past models, thus making a significant theoretical and methodological contribution.

4.2.2 Citation Tracking Approach

A separate aim of many diffusion studies is to identify the level of interest in a particular topic or idea over time. The bibliometric technique of article counting has generally been considered a reliable approach for analyzing the published literature to trace the development of an innovation (Abrahamson & Fairchild, 1999; Clark, 2004). It involves

capturing annual citation counts to determine how many academic articles have focused on a given concept over time; the time-series data can be charted into a lifecycle and compared to the most well-known lifecycle shape (the S curve depicted in Figure 2.1) (Abrahamson & Fairchild, 1999; Ponzi & Koenig, 2002). Bibliographic records are a relatively objective indicator measuring discourse popularity surrounding an idea. Scholars have thus adopted this method, assuming that the higher the citation counts, the larger the volume of discussion/interest in a particular topic (Ponzi & Koenig, 2002; Thomas, 1999). Notably, *"citation is a cumulative number and the duration is very long, so we can map the entire diffusion process of the innovation"* (Zhai et al., 2017, 370).

While this method has been widely used (Small, 1978; Zhai et al., 2017; Zhang et al., 2013) it has well-known limitations (Clark, 2004; MacRoberts & MacRoberts, 1996). Articles relate to academic, rather than practitioner, interest. They are a lagging indicator since academics often work behind the practitioner curve and studies take time to be published and cited. Also, articles do not appropriately capture changing sentiments – that is, attitudes, thoughts or judgements about the reform. Thus, analyzing citation counts in isolation could result in misleading opinions about the development of an innovation, necessitating the inclusion and analysis of other forms of data.

4.3 Rationale for approaches used in this thesis and rejection of alternate methods

Methodologically, in this thesis, I seek to demonstrate the limitations of past approaches in explaining the rapid growth of early-stage public sector innovations, and in doing so, to test and advance alternative approaches. While an intuitive first-step in understanding why SIBs have diffused throughout various public sectors, was to analyse the existing academic and grey literature, using systematic review methods, it was the social

media analysis that proved to be particularly novel. Within this, a particularly crucial component, was to show that analyzing sentiments can provide valuable insights which assist with understanding institutionalization. To date, sentiment analysis has rarely been incorporated into empirical institutional studies. Social network analysis has been drawn upon in various studies analyzing the diffusion and/or institutionalization of social innovations (e.g. Kolleck, 2013). However, studies complementing sentiment analysis with sociological theories are rare. Thus, further empirical research is required to understand the extent to which sentiments can assist with identifying instrumental and/or institutional drivers of adoption. This gap is gradually being addressed, exemplified by the fact that some recent studies have highlighted the usefulness of conducting sentiment analysis to understand perceptions expressed in social media in the context of the diffusion of innovation (Al-Razgan et al., 2021). For instance, Bolici et al., (2020, p. 1) note that measuring the sentiments contained in Tweets *"provides an interesting overview of the individuals attitudes towards the awareness or the adoption of new technologies."* Regardless, there is a need for further advancements in terms of combining social media analysis with sociological theories, particularly in the context of early-stage public sector innovation.

Using the Twitter API allowed for a unique dataset to be built, which addressed the various elements of Rogers's theory. While a small number of SIB studies have drawn on the institutional literature, they have not combined this theoretical lens with social media data. Thus, no other studies have used the sentiments contained in social media posts and/or messages as an indicator of SIBs policy diffusion. Given the severe data limitations and evaluation difficulties in this field, it was crucial to maximize existing data as much as possible, including social media data. Sentiment analysis could also be applied to other (non-

Twitter) data; this should be explored further in future SIB studies, given the availability and richness of non-quantitative data.

In addition to understanding the rationale for using the selected methods, it is integral to consider why particular approaches were overlooked. While approaches such as citation tracking and regression analysis (similar to Tolbert and Zucker's 1983 study) were dismissed due to being unsuitable in the SIBs context (as well as due to the well-documented limitations associated with these methods), others, namely qualitative interviews were not included, primarily for pragmatic reasons. The scope of this thesis originally included a qualitative investigation involving semi-structured interviews with both adopters (i.e. policymakers and other practitioners) and non-adopters of the SIB contracting model. However, given the richness of the systematic review and social media analysis chapters, it was subsequently deemed that the interviews were not needed for this thesis. An enormous amount of existing qualitative data (from academic studies, grey literature and social media data) were retrieved and analyzed, thus meaning that 'saturation' of qualitative findings had already been achieved, without requiring the collection of additional qualitative data. Additionally, the social network analysis revealed that stakeholders such as policymakers and other SIB practitioners (e.g. consultants, investors, service providers etc) were represented and influential in the Twitter dataset (although to differying extents). Thus, the perceptions of these stakeholders were captured and analyzed, in the social media chapters. Nonetheless, it is important to note that a qualitative study (featuring a sample of policymakers and practitioners) could have further reinforced the findings and also assisted with overcoming methodological limitations and gaps in the diffusion of innovations literature. For example, many of the qualitative studies included in the qualitative systematic review did not explicitly aim to identify motivations for adopting SIBs; thus, including a qualitative study would have

permitted for more tailored aims and interview questions. More broadly, early diffusion studies placed an over-reliance on models of diffusion which are too rationalistic (e.g. assuming that all adopters view an innovation in a particular way); thus making it important to conduct additional process research which involves investigating the broader context in which the innovation diffuses (Belasco, 1989; Rogers, 1995). While not included in this thesis, future qualitative work will be undertaken to further test and refine the overarching findings and to assist with translating the findings into practical tools that can be of use for stakeholders such as policymakers.

4.4 Key Methodological Components

The various limitations (of applying one model or approach) highlighted above made it particularly crucial to use various types of evidence. Thus, taking a consilience approach, used a diverse array of data and tests (Hood & Dixon, 2010; Laudan, 1971; Whewell, 1840). Hood and Dixon (2010, p. i286) liken the approach to the *"accumulation of pieces of circumstantial evidence in a legal case that together can meet the standards of civil or criminal proof."* It was particularly integral for this thesis, as the SIB concept is an example of a complex intervention; and measuring the interplay between instrumental and institutional motives is particularly challenging. It was impossible to identify one definitive test or data source which could answer my research question. Nevertheless, consistent results, drawn from different data sources and methods were used to support or refute the hypotheses more convincingly than relying *"only on a single source or very closely related phenomena"* (Hood & Dixon 2010, p. i286).

The 'consilience' approach differs from standard social-scientific inference methods, which aim to establish the link between perceived cause and effect, often using RCTs

(Hernán, 2004; Glass et al., 2013; Zwahlen & Salanti, 2018). Counterfactuals are a crucial component of a causal inference framework; 'counterfactual outcome' is used to describe a hypothetical outcome which is not observed (Rubin, 1974; Hernán, 2004). The inclusion of a counterfactual allows researchers to mathematically define a causal effect. Thus, causal inference commonly involves estimating the average treatment effect of an intervention (Moser et l., 2020). While it may intuitively seem that the consilience approach is of a lower scientific standard than causal inference, a more nuanced view is required, noting that researchers have called for a pluralistic approach to causal inference. For instance, Vandenbroucke et al., (2016, p. 1776) noted that *"it would be of concern if the theory and practice of the complete field of epidemiology were to become restricted to this single approach to causal inference".* Similarly, Claveau (2011, p. 233) urged researchers to take advantage of approaches that seek *"variety of evidence – understood specifically as using multiple means of determination to robustly estimate the same causal effect."* Thus, the consilience approach is consistent with the gradual shift away from solely relying on RCTs for assessing causality. It should therefore be viewed as complementary to the causal inference approach rather than of a lower standard.

Table 4.1 depicts which of the main drivers each part of the thesis primarily focused on exploring. I hypothesized that the desire among governments to secure legitimacy among valued stakeholders through adopting a novel idea with private-sector undertones has driven the rapid and widespread adoption of SIBs. Regardless, it was still essential to explore, and discount other motives, such as instrumental drivers. Thus, data from each section were triangulated, as evidence from each stand-alone section produced ambiguous evidence with multiple (possibly competing) explanations.

Table 4.1: Contribution of Each Empirical Component

Study details				Evidence primarily testing for		
	DPhil Component	**Research Question/s**	**Data source/s**	**Instrumental Drivers**	**Institutional Drivers**	**Other Possible Drivers**
Systematic reviews	**Effectiveness systematic review**	What are the effects of any intervention funded by a SIB, compared to the absence of intervention or to alternative interventions, on any social outcomes?	Published experimental and quasi-experimental evaluation reports for SIB-funded programs.	X		
	Qualitative systematic review	What are the views of SIB practitioners regarding motivations for adopting SIBs, and what does this reveal about instrumental and/or institutional drivers?	Studies that use qualitative data to explore SIB practitioner perspectives about the outcomes-based contract.		X	
Social media analysis	**Analysis of Social Media Data – statistical analysis, sentiment analysis, social network analysis**	How have sentiments regarding SIBs changed over time on Twitter, and to what extent do the sentiments expressed in the tweets express emerging (negative) research findings? Which stakeholders have played a central role in communicating and promoting the SIB idea in online forums such as Twitter?	Dataset of all Tweets posted between 2010 – 2020 that used a relevant SIB hashtag, collected using Twitter's Full Archive Application Programming Interface (API). Also, data from Wikipedia Page Views, Google Trends, Google Ngram Viewer and Reddit – for comparisons.	X	X	X
	Analysis of Social Media Data – framework analysis	What discourses can be identified in the highly shared tweets, and to what extent have particular discourses dominated/others been drowned out? What persuasive techniques have been used by SIB adopters to frame the SIB funding model as a 'legitimate' reform to pursue over time?	Twitter dataset, as described above – although subsets of this (most highly retweeted tweets and a random sample).	X	X	X

In addition to the specific data sources outlined in Table 4.1, I referred to several of the GO Lab's International Network for Data on Impact and Government Outcomes (INDIGO) datasets throughout this thesis. Thus, I will now provide a brief rationale for using

this data source, and I will also outline the strengths and weaknesses of it. INDIGO is a data collaborative which includes various community activities, a system for sharing SIBs data and various datasets that are publicly available on the GO Lab website. In this thesis, I primarily used their Impact Bond dataset, although I also briefly referred to their Pipeline Dataset and their Network Evolution data.

INDIGO provides the most complete, open dataset on SIBs (Impact Bonds Dataset), which is widely drawn upon, to understand the growth and current state of the SIBs landscape. Importantly, INDIGO datasets are publicly available and can thus be easily accessed. Additionally, the GO Lab (a research and policy centre based at the University of Oxford) have devoted (and continue to devote) extensive time and resources to updating the datasets. As the GO Lab are regarded as a credible global knowledge hub, the INDIGO datasets are perceived as reliable sources of information, despite their shortcomings. Alternative SIB datasets have been made available at different points in time (e.g. by organizations such as Instiglio, Social Finance etc). However, these no longer appear to be available, for example, with Social Finance now redirecting people to the GO Lab Impact Bond dataset, from their own website. Additionally,the Brookings Institute have developed and maintained a ‘comprehensive’ database on the global SIBs market since 2014; yet, they only present monthly snapshops from the database, rather than making their entire database publicly available (and multiple requests for access to the database, to assist with this DPhil thesis, were unfruitful). Thus, the only alternative to using the GO Lab’s data, would have been to build an additional dataset, although this was deemed to be undesirable, given the time consuming nature of doing so; and the fact that it would have duplicated work already undertaken by the GO Lab.

INDIGO relies upon SIB projects from across the world volunteering their data. Consequently, there are obvious limitations associated with relying upon INDIGO data. Firstly, the data contained in the INDIGO Impact Bonds dataset are incomplete, with many projects having been launched but not added to the dataset (e.g. in NSW, the Silver Chain SIB and Home and Healthy SIB). Additionally, many project listings contain missing information (particularly sensitive information, such as investor returns); and timely updates about listed projects are not always available. Secondly, there are inconsistencies in the information contained in the Impact Bond dataset, for example, with discrepancies between the headline figures reported in the graphs and those obtained by manually calculating the aggregated results from the information contained on the individual project pages. Third, the Impact Bond dataset contains limited information on project terminations (e,.g. with the 'stage of development' filter only including 'contracted'. 'implementation', 'service delivery complete', and 'complete' stages), thus making it difficult to assess the extent to which SIBs have been abandoned, as well as adopted. Finally, while INDIGO has a process for processing information and they attempt to attribute data to named sources, they do not audit, cross-check or very all of the information provided to them; the quality of the data on an individual project is thus highly dependant on the skills and resources available to the relevant project. Nonetheless, INDIGO provides an important source of information on SIBs and the above-mentioned limitations are outweighed by various strengths, particularly the fact that INDIGO provides the most comprehensive source of information on SIB projects.

4.4.1 Effectiveness systematic review

Overview of method and justificication for use

Systematic reviews address a specific research question by collating evidence fitting pre-specified eligibility criteria (Higgins et al., 2019). By using systematic methods, they minimize bias and are thus considered the 'gold standard' for identifying and synthesizing the best available evidence regarding a specific question (Liberati et al., 2009; Munn et al., 2018). The systematic review methodology first rose to prominence in the 1980s and 1990s with the development of the EBM movement (Evidence-Based Medicine Working Group, 1992; Sackett et al., 1996). This movement responded to the concern that medical interventions were not always based on robust evidence of effects. Consequently, systematic reviews have traditionally been used to assess the effectiveness of health interventions, through synthesizing the results of RCT studies (using meta-analysis where possible) (Munn et al., 2014; Pearson, 2004). In this context, effectiveness refers to *"the extent to which an intervention, when used appropriately, achieves the intended effect"* (Munn et al., 2018, p. 2; Pearson et al., 2005). Since the election of the 1997 New Labour government in the UK, there have been calls for the broader use of EBPM. This has included calls for the use of systematic reviews in social policy areas such as social care, education, and crime (Davies et al., 2000). While systematic reviews are now used in diverse policy contexts, and various review methods have emerged (e.g. mixed methods), traditional effectiveness reviews remain common (Munn et al., 2018).

As this DPhil thesis addresses questions that go beyond 'what works', it was initially envisaged that a mixed-methods approach would be used for the systematic review component (Hannes & Lockwood, 2012). This approach was developed by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre). It combines the findings of qualitative and quantitative studies within a single systematic review, thus addressing complementary or overlapping research questions (Campbell et al., 2000; Medical

Research Council, 2000; Thomas & Harden, 2008). As with a standard systematic review, the review question should dictate the suitability of a mixed methods approach (Lizarondo et al., 2022). In the SIBs context, formal methods for mixed-methods reviews (outlined by the EPPI Centre) would be unlikely to yield useful results due to the low quality and heterogeneity of the effectiveness findings. Additionally, in this instance, the quantitative and qualitative research questions were not complementary to the extent usually expected in a mixed methods review. When a mixed-methods approach was originally planned, the qualitative question was linked to the effectiveness question, as it was focused on identifying the strengths, critiques and 'active ingredients' of the SIB approach (which could help to explain the effectiveness data). However, the qualitative research question later shifted to one centred around the pillars of legitimacy and identifying motivations for adoption; this deviated significantly from the quantitative review, making it difficult to envisage how triangulation would occur. Consequently, two stand-alone systematic reviews were conducted; the findings were juxtaposed and compared in a more informal manner where possible, for instance, in the closing chapter of the thesis. As my overarching research question seeks to identify what has driven the diffusion of SIBs, the natural starting point, was to synthesise the available quantitative evidence, to understand how evidence of effectiveness has shifted over time. Below, I provide a brief overview of the effectiveness review methods used in this thesis.

Summary of application in this thesis

The effectiveness review helped to understand whether SIBs are efficient and at what point in time evidence of efficiency did or did not emerge. This was crucial for understanding the extent to which “instrumental” motives (i.e. related to furthering organizational objectives throughthe rational selection of management tools and techniques) can partly or wholly explain SIB adoption. Using the PICO framework (Melnyk & Fineout-Overholt, 2015), the

research question asked: *"What are the effects of any intervention funded by a SIB, compared to the absence of intervention or to alternative interventions, on any social outcomes?"* The hierarchical exclusion criteria noted that to be included in the review, a study had to: 1) used an experimental or quasi-experimental design; 2) focus on a SIB-funded intervention; 3) include primary date about the effect of the SIB funding model and/or the intervention on service provision; 4) be published in English; and 5) be retrievable.

It was important to search a diverse array of databases, to retrieve a broad range of published and unpublished literature. Thus, in total, 12 electronic databases were searched. Additionally, a comprehensive grey literature search was undertaken (Appendix B), as well as hand searching of 48 journals (Appendix C). I contacted authors in this field (Appendix D) and searched the reference lists for relevant studies and reviews. I combined the studies identified from electronic searches and additional searching (grey literature, hand searching, snowballing, contacting experts) in Covidence (systematic review software) and removed duplicates. I then screened all articles on title and abstract, excluding articles that did not meet one or more components of the hierarchical inclusion criteria. I reviewed the full-text of remaining articles and assessed them for eligibility using the predetermined inclusion criteria.

For data extraction, I used an adapted version of Cochrane's Effective Practice and Organisation of Care (EPOC) data collection form (Appendix F). I critically appraised the quality of each included study using a tool designed specifically for quantitative research. The risk of bias in randomized studies was assessed using Version 2 of the Cochrane risk-of-bias tool for randomized trials (RoB 2) (Sterne et al., 2019). For non-randomized studies, I used the ROBINS-I tool for assessing risk of bias in non-randomized studies of interventions

(Sterne et al., 2016). To assess broader considerations related to study quality, I used a standardized quality assessment tool was used (Appendix G), specifically, a modified version of the Hamilton tool ("Quality Assessment Tool for Quantitative Studies" – Effective Public Health Practice Project 2007) (EPHPP, 1998).

In regards to data analysis, it was not possible to conduct meta-analysis, due to the scarcity of experimental and quasi-experimental studies containing relevant empirical data and significant heterogeneity of studies. Thus, the results of the effectiveness review were presented using narrative synthesis. This approach is consistent with that used in many other studies, where similar data limitations have been encountered (e.g. Lorenc et al., 2014). Where available, I analyzed effect sizes as change scores (i.e differences between pre and post) and relative change scores (i.e differences between change scores in the intervention and control groups). I standardized ordinal outcomes, such as Likert-type scales, to 0-100. In order to summarize effect sizes across multiple distinct outcomes and/or studies, I calculated median change scores, thus providing an indicative measure of effect size for each study. I synthesized the available outcome data separately for controlled and uncontrolled studies, thus allowing for any divergence between these two groups to be identified.

4.4.2 Qualitative systematic review

Overview of method and justificication for use

Our knowledge of the factors that have driven the continued uptake of SIBs is limited, and few attempts have been made to systematically analyze the qualitative literature to better understand this. In fact, systematic reviews of public management reform processes are scarce, even beyond the case of SIBs. It is unclear whether the uptake of SIBs has been largely driven by instrumental or institutional motives (or both), and how the motives have

differed at different stages of the adoption cycle. Thus, a qualitative systematic review was also conducted to build on the findings of the effectiveness review. While policymakers initially aimed to assess intervention effectiveness ("what works?"), they soon realized the importance of expanding this to include *"what works – for whom – and under what circumstances"*, along with questions about health needs, acceptability and appropriateness (Hannes & Lockwood, 2012). Answering these diverse questions requires qualitative research, particularly research exploring the perspectives and experiences of the relevant population groups. Thus, the standard systematic review model (used to address intervention effectiveness questions) was adapted by incorporating methods already established for data analysis in qualitative research (Hannes & Lockwood, 2012).

A qualitative methodology is often considered appropriate when questions such as how, why and what are central; and when the topic of interest is emerging (Creswell, 1998; Bowling, 2002). Qualitative research involves *"the systematic collection, ordering, description and interpretation of textual data generated from talk, observation or documentation"* (Kitto et al., 2008). Qualitative systematic reviews analyze human experiences and cultural and social phenomena (Munn et al., 2018). This type of review can be helpful for understanding why an intervention has not been adopted despite evidence of its effectiveness and vice versa (Hannes et al., 2010; Lockwood et al., 2015; Munn et al., 2018; Pearson, 2004).

While qualitative research is now recognized as a critical part of intervention research, it has not achieved the same level of support as some quantitative methods (e.g. RCTs); and is often overlooked (Dixon-Woods et al., 2006; Fraser et al., 2009; Greenhalgh, 2010; Oakley, 2002). The existence of 'paradigm wars', in which researchers attempt to

discredit the validity of each other's work and methods, contributes to this (Davies & Nutley, 2002). Traditional systematic review methods have elevated the results of experimental research to a superior position in terms of quality and applicability (Pearson 2004). Critiques of this note that although traditional scientific methods play a crucial role in developing knowledge and technology, it is vital to challenge the dominance of traditional science, which emphasizes objectivity (Fitzgerald, 1995). Recognising this, I built on the effectiveness systematic review (Chapter 5) with a qualitative systematic review (Chapter 6). Below, I provide a brief overview of the qualitative systematic review method that I used in this thesis.

Summary of application in this thesis

In brief, published studies were eligible for inclusion in this thematic synthesis (Chapter 6) if they used qualitative methods to study the motivations underpinning the adoption of SIBs. Very few studies have an explicit objective of understanding why SIBs have been pursued; and motivations can be difficult to identify (and are not always explicit). Thus, if studies lookede at factors such as the benefits, critiques and/or active ingredients of SIBs (and in doing so, broadly reflectd on motivations), they were deemed to be of relevance and eligible for inclusion. As with the effectiveness review (Chapter 5), the included studies were identified through exhaustive searching of 12 electronic databases, comprehensive searching of the grey literature (Appendix B), hand-searching of 48 journals (Appendix C), contacting authors in this field (Appendix D) and screening the reference lists of included studies in relevant previous reviews (e,g. Broccardo et al., 2019).

After identifying potentially relevant studies using the above methods, all studies were merged (using “Covidence” software) and duplicates were removed. Studies wereinitially be screened on title and abstract; any obviously irrelevant studies were removed.

Full-text review was undertaken for the remaining studies. For pragmatic reasons, articles were not double screened. Data were extracted on factors such as study aims, study settings, study sample, data collection, and key findings. Four distinct types of data sampling were applied to the studies selected from the overall sampling frame, including: 1) intensity sampling (i.e. selecting studies of high quality and that contain rich data on motivations for SIB adoption; 2) maximum variation sampling (i.e. including studies with unique or diverse characteristics that are not identified through the intensity sampling); 3) confirming and disconfirming cases (i.e. identifying studies that confirm or disprove the patterns already identified); and 4) politically important cases (i.e. ensuring the sample includes studies focused on 'hot', politically sensitive examples).

While the aim of the thematic synthesis is not to determine the size of an effect, critical appraisal was integral for assessing confidence in the findings and for conducting sensitivity testing. Thus, all included studies were subject to appropriate critical appraisal, using the Critical Appraisal Skills Programme (CASP) qualitative appraisal tool (CASP, 2018). I also assessed the contribution of each study to the synthesis (as recommended by Rees et al., 2009); and used the Confidence in the Evidence from Reviews of Qualitative research (GRADE-CERQual) approach for qualitative syntheses (Guyatt et al., 2011) to identify the level of confidence to place in each finding (Lewin et al., 2015).

In regards to the thematic synthesis process, I closely followed the procedures outlined by Thomas & Harden (2008). Firstly, to identify descriptive themes, all data were coded line-by-line. I assigned at least a single code (and in many cases, multiple codes) to all sections or fragments of text (e.g., sentences) in order to describe their content and meaning. Secondly, as the descriptive codes began to accumulate, I organized them hierarchically, into

categories falling under the well-documented legitimacy pillars. I reconsidered the codes in an iterative manner, as required, to incorporate newly identified content. Thirdly, I developed analytical themes by reading and re-reading all descriptively coded data and comparing and contrasting the descriptive themes identified across studies. Through this process, I developed a set of 'factors' that influence the uptake of SIBs by stakeholders such as policymakers.

4.4.3 Quantitative and Qualitative Analysis of Social Media Data

Overview of method and justificication for use

A mixed-methods analysis of social media data is an essential component of this thesis as legitimacy is a continually unfolding social process (Greenwood et al., 2017). Using multi-year social media data allowed for an analysis of how this process has unfolded over time. In addition, the information dissemination and information-seeking process has become increasingly digitalized, leading to new ways of measuring public attention and awareness (Eysenbach, 2009; Rimer & Glanz, 2005; Jacobs et al., 2017). In a digital era, it is important to make full use of the available data, including real-time data created online through citizens and organizations interacting (Mergel et al., 2016). There are many benefits to using data produced from digital platforms; for instance, it is possible to track interactions on a much larger scale and to understand the opinions and preferences of subjects without directly asking them questions (Chang et al., 2014). Often termed 'trace data', Howison et al., (2011, p. 769) note that there are three differences to traditional research data, including: 1) it is 'found', rather than being directly produced for research; 2) it is event-based data, concerned with single interactions; and 3) the 'events' have a temporal dimension meaning that trace data are naturally longitudinal data. 'Trace' data can assist in the study of innovation diffusion, specifically with identifying legitimacy versus efficiency drivers through capturing the thoughts and perceptions of individuals and organizations, regarding an innovation,

across time. With these benefits in mind, an increasing number of researchers are now using data from digital platforms such as Twitter, Facebook and Instagram to better understand the behaviours and opinions of individuals and populations (Hu, 2019).

The data available through social networking sites (SNSs) such as Twitter are produced at a fast rate, allowing for large and unique datasets to be built. These datasets can make a valuable contribution to the field of public policy (Hu, 2019). SNSs are web-based platforms that enable users to create profiles, post information, view and/or react to the information posted by other users and make connections to others (Boyd & Ellison, 2008: 211). The interaction feature distinguishes SNSs from other web platforms (Medaglia & Zhang, 2017). Twitter is *"a microblogging and social networking service on which users post and interact with messages"* (i.e. 'tweets') (Erskine & Hendricks, 2021, p. 2). It is among the most popular SNSs with 330 million monthly active users as at May 2020 (when the company discontinued reporting of this statistic) (Statista, 2021a). Twitter data are well-recognized by many researchers as offering significant potential for unpacking new social relations and phenomena. Thus, Twitter research is well supported by various funding organizations (Sinnenberg et al., 2017; Tinati et al., 2014). Regardless, social media data are yet to be incorporated into academic studies focused on SIBs.

Including social media data (particularly Twitter) is critical for incorporating otherwise overlooked perspectives, noting that stakeholders such as consultants are often very active in the early stages of the management fashion life cycle, and much of the discussion occurs online. This is exemplified by enthusiastic social media posts emphasizing the merits of a new concept in the early stages of diffusion. Conversely, academics often play a crucial due diligence role in regards to management fashions. However, due to slow publishing and

review processes, academic discourse tends to surface later in the cycle of management fashions (Heusinkveld, 2004, p. 27). Thus, merely analyzing academic discourse would be insufficient for addressing the objectives of this thesis. In this thesis, the inclusion of social media data allowed for an understanding of the diffusion of SIBS and the changing tenor of public, practitioner, expert and consultant "chatter" about this innovation over time.

Summary of application in this thesis

The dataset used for the social media analysis component of this thesis was built using Twitter's Full Archive Search API. I captured tweets posted between 2010 - 2020 which used a relevant SIBs hashtag (e.g. #socialimpactbond, #socialimpactbonds, #socialbenefitbond, #socialbenefitbonds). Tweets which mentioned SIBs but failed to use a relevant hashtag were not retrieved. As the dataset spans multiple years, it provided valuable insights into how sentiments towards SIBs have differed over time, including how (if at all) the changing sentiments appear to have impacted the uptake of SIBs and vice versa (noting that uptake of SIBs may, conversely, have affected changing sentiment).

In Chapter 7, I firstly calculated descriptive statistics for the entire dataset to identify patterns in factors such as levels of interest over time. I compared the patterns identified in the Twitter data to those identified in other online sources of data, such as Reddit, Google Trends, Google Ngram Viewer and Wikipedia Page Views. Regognizing the importance of understanding the attitudes and emotions contained in tweets, rather than just the volume of tweets, I then applied pre-trained sentiment models to the whole dataset to analyze how sentiments have differed between 2010 to 2020. A sentence-level sentiment score was calculated for each tweet based on a pre-trained sentiment analysis model (initially "sentimentr" package in R). The scores obtained using 'SentimentR' were then compared to

those calculated from other well-known packages, namely RSentiment, Sentiment Analysis, Syuzhet and MeanR. I also compared the sentiments calculated using the automated packages, to those identified, through manual hand-coding of the entire dataset. While an alternate approach would have been to train a sentiment analysis model based on a manually annotated corpora of SIB specific content, this would have evidently required significant human effort and caused considerable time delays and may not have necessarily produced more accurate results. Thus, pre-trained packages were used when analysing the entire dataset, although the patterns dervived from this analysis were compared to those identified from manual sentiment analysis (i.e. reading each tweet and coding it as having a positive, neutral or negative attitude towards SIBs). As a final component of this chapter, I conducted social network analysis in R to visualize the SIB network and to understand who has been particularly influential in spreading the idea. This involved conducting Social Network Analysis (SNA) using the igraph R package. I identified the most influential accounts, according to key centrality measures, and presented the results in table format.

In Chapter 8, I used rich qualitative methods to build on the broader, largely statistical findings of the prior chapter. This analysis involved two subsets of the larger dataset. Firstly, I used R (statistical software) to construct a random (10%) sample of the original tweets contained in the dataset; this included both adopters and non-adopters of SIBs, from different time points and different jurisdictions. I also created a purposive sample consisting of the most highly retweeted tweets, which was used for the discourse analysis. I used framework analysis to analyse both of these subsets of the data. This involved the following five stages: 1) data familiarization (i.e. reviewed a subset of the tweets and developed initial ideas); 2) framework identification (i.e. selected a framework for the analysis and resulting interpretation, which draws on a priori assumptions and concepts); 3) indexing (i.e.

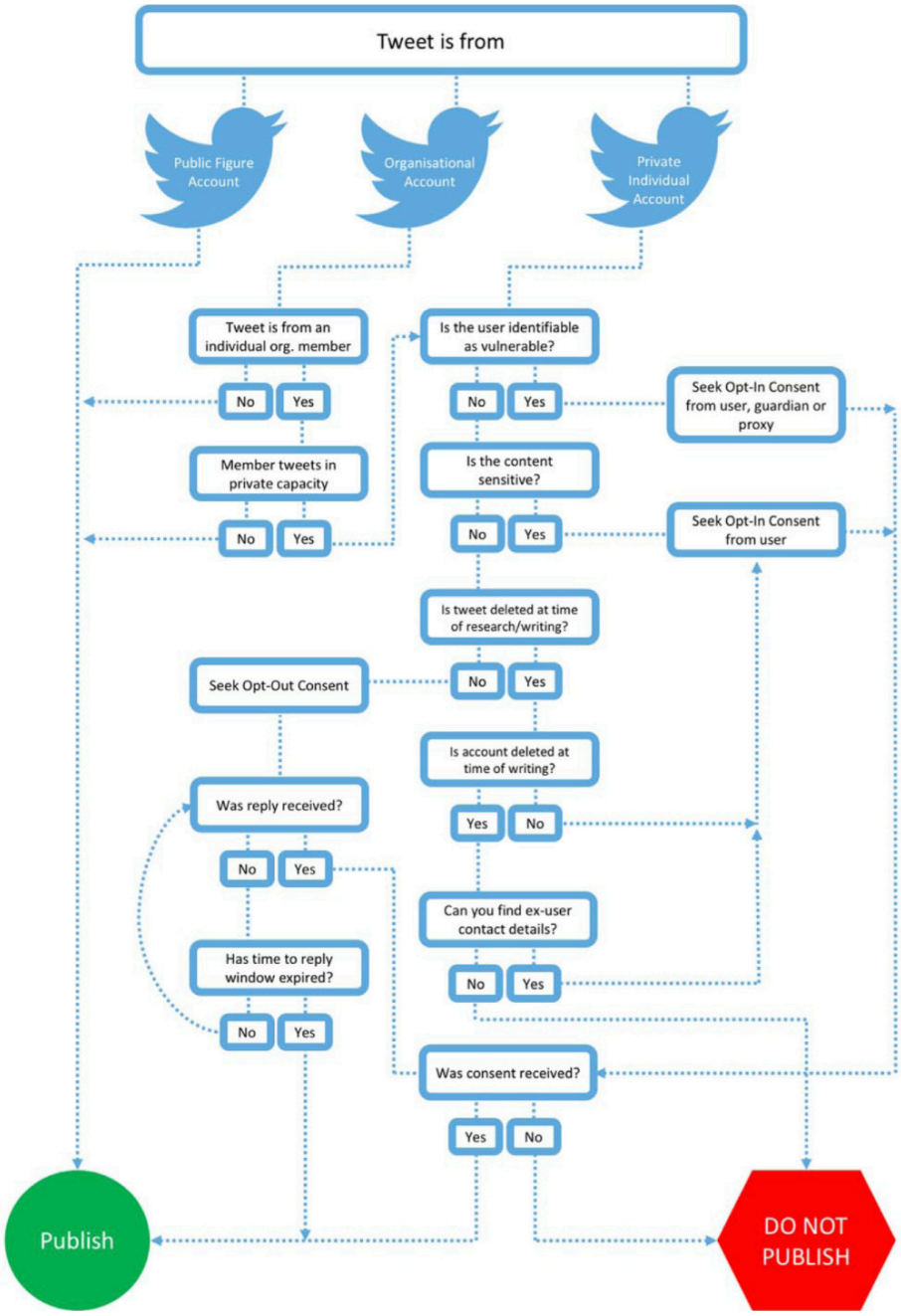
systematically applied the framework to all of the data); 4) charting (i.e. ordered and indexed the data in a matrix form); and 5) mapping and interpretation (i.e. combined the key learnings from the earlier steps) (Ritchie & Spencer, 1994). For the random sample, I operationalized the 'acceptance factor' items outlined by Hood and Jackson (1991), and for the purposive sample of highly retweeted tweets, I used (and built upon) a discourse framework developed in an existing SIBs-focused study (Ormiston et al., 2020). I discuss the application of these frameworks in further detail, in Chapter 8.

In addition to applying the above frameworks, I created a scale that measures the extent to which each tweeter has a conflict of interest (COI) and considered how this has influenced factors such as sentiments expressed. It is well-recognized that COI can arise in various ways in evaluation studies; an important distinction needs to be made between financial COI (i.e. possible material gains) and non-financial COI (i.e. other benefits such as enhanced reputation) (Eisner & Humphreys, 2012; Eisner et al., 2015; Resnik, 2000). While there are many stakeholders who stand to gain from the success of the SIB concept, no COI instruments have been developed or tested in this context. As indirect benefits can be difficult to operationalize, I developed a COI framework which, consistent with past studies, involved collecting information that is primarily related to identifying a potential financial COI (Eisner & Humphreys, 2012; Krimsky et al., 1998; Friedman et al., 2004). The specific framework is outlined in Chapter 8; broadly, it assumed that a COI is more likely if the tweet author (organization or individual) has been involved with one or more SIB contracts; and if the category of involvement was one where there could be a financial gain (e.g. investor, intermediary etc).

4.5 Ethical considerations

When analyzing Twitter data, several ethical issues need to be considered (Pagoto & Nebeker, 2019; Webb et al., 2017; Williams et al., 2017). This study adopted an ethics framework proposed by Williams et al., (2017) to consider ethical risks and mitigate any harm. This involved collecting the data without explicit consent but seeking informed consent for any directly quoted content. For situations in which the content included in the tweet is not sensitive, and the user is not vulnerable, opt-out consent was considered sufficient (Williams et al., 2017). Figure 4.3 presents the specific process used in this study when considering whether to include a verbatim quote.

Figure 4.3: Flow Chart for Risk Assessment - Publishing Verbatim Quotes



Source: Williams et al., (2017, p. 1163)

4.6 Conclusion

In this chapter, I have outlined the methodological approach that was used to determine what explains the rapid and widespread growth of SIBs. I described how the

methods appropriately addressed the main hypotheses outlined in Chapter 3 of this thesis. Notably, I conducted an effectiveness systematic review, followed by a qualitative systematic review, to understand what the changing landscape of evidence reveals about possible instrumental and/or institutional motivations for adoption. The systematic review chapters assisted with identifying how knowledge about the effectiveness (or otherwise) of SIBs has shifted over time, leading to changing levels of uncertainty. I then analyzed the available social media data (using a combination of statistical analysis, sentiment analysis, network analysis and framework analysis) to consider what this data source highlights about how the idea has been socially constructed to be legitimate, by particular actors, over time. I drew together evidence from multiple data sources and methods, using a form of 'consilience' in collecting and analyzing data. Having outlined the methodological approach, the next portion of this thesis presents the results of the empirical research.

Part Two: Empircal Research

Chapter 5: A Systematic Review of Effectiveness

5.1 Introduction

This chapter presents a systematic review of the literature investigating the effectiveness of the SIB funding model on service delivery. It explores the changing state of academic knowledge surrounding the benefits of SIBs during 2010-22. This is important given that I am predicting that the mix of instrumental and institutional, or efficiency and legitimacy, pressures for public management reform are influenced by the level of uncertainty regarding intervention effectiveness. Specifically, I argue that greater and/or ongoing uncertainty contributes to decreasing efficiency and increasing legitimacy drivers over time. In the context of SIBs, it is possible that early policymakers may have pursued the reform for efficiency reasons, despite a lack of evidence, given that some level of uncertainty is expected in the initial stages of adopting an innovation. However, as uncertainty has remained due to a lack of evidence and some negative evaluation findings have accrued over time, it seems less likely that middle- to later-stage adopters would have also pursued the reform for technical gains. When results cannot be properly observed and measured, the 'institutionalization dynamic' does not allow for an idea to be quickly abandoned due to disappointment (Rao et al., 2001). Thus, I argue that the ongoing uncertainty surrounding the effectiveness of SIBs has prolonged interest in the idea and influenced the mix of efficiency and legitimacy pressures facing policymakers.

Although actual usage of SIBs has declined over time (GO Lab, 2022a), general enthusiasm remains among political leaders in Britain and other countries such as Portugal, Japan and Russia. Several grey literature papers have heavily endorsed the reform and claimed it has been overwhelmingly successful. For instance, Stanworth & Hickman (2022) noted that the funding model had produced a ten-fold return for the government. However,

reports such as these lack rigour and have been publicly criticized (Ball, 2022; Patton, 2022). The majority of the existing evaluations lack a robust control group, making it difficult to identify a causal link between the SIB funding model and the outcomes resulting from the SIB-funded intervention. Few SIB-focused systematic reviews have been conducted. Existing reviews have not quantitatively analyzed how the SIB funding model affects service provision (i.e 'SIB effect'). This effectiveness review, therefore, builds on prior studies by using a broader research question and more comprehensive search strategy and including evidence from all jurisdictions and policy areas (subject to the full text being available in English). By searching a larger number of academic databases and grey literature sources, this study has identified SIB projects and evaluations that have yet to be incorporated into systematic reviews. In this chapter, I find that the early to middle years were categorized by uncertainty, with a complete lack of quantitative evidence from 2010 – 2013. Some evidence emerged from 2014 – 2017, although predominantly with null results; the only study with positive outcomes used a non-randomized design (DCLG, 2017). From 2018 – 2022, the evidence base was strengthened, with some positive results emerging from randomized studies (Cunningham et al., 2021; Gallardo et al., 2021; ID Insight, 2018; Raven et al., 2020; Hunter et al., 2022). However, uncertainty remains regarding the effectiveness of the funding model itself as existing evaluations only focus on individual scheme-level outcomes, thus making it impossible to aggregate evidence at a higher level, at this point in time.

5.2 Overview of methods

This study is a systematic review of empirical literature related to the effectiveness of SIB funding programs. It provides a starting point for understanding how the landscape of SIBs evidence has developed over time. Specifically, the effectiveness review helps to understand whether SIBs are efficient, at what point in time evidence of such efficiency did

or did not become available to practitioners, and thus to what extent “instrumental” motives concerned with furthering organizational objectives through the rational selection of management tools and techniques can explain SIB adoption.

The research question for the effectiveness review originally asked: *“what is the effectiveness of the SIB funding model on service provision, as compared to other types of commissioning models?”* However, an early (2018) version of this systematic review (presented at the Transfer of Status milestone) showed that this would result in an ‘empty review’ (i.e. no studies eligible for inclusion) and that potentially relevant evidence (studies not including a SIB comparator) would be completely overlooked (Yaffe et al., 2012). While ‘empty’ reviews can play an important role (e.g. by highlighting evidence gaps), it was deemed more important to find a way to include the available studies that have incorporated a sufficiently robust study design. Thus, when updating the effectiveness review, the criteria were altered to remove the requirement for a comparison to non-SIB commmissioning models. Table 5.1 shows each element included in the research question using the 'Population, Intervention, Control, and Outcomes (PICO) format. This is a well-known strategy for forming a research question, which facilitates the development of questions intended to assess the effectiveness of interventions (Melnyk & Fineout-Overholt, 2015).

Table 5.1: PICO Research Question

Element	
Population	SIB practitioners who have been (or are currently) involved in delivering a social intervention that has been funded using a SIB structure
Intervention	Use of the SIB funding model to fund a social intervention (i.e. upfront capital from private investors)
Comparison	Compared to the absence of intervention or to alternative interventions (i.e. study design includes some form of a control group)

Outcome	Any social and/or environmental outcomes. This could include outcomes at the level of the funding model or service level outcomes.

Search strategy

Inclusion and exclusion criteria

An explicit inclusion and exclusion criteria (Table 5.2) was established and applied to reduce bias and to distinguish the systematic review from a narrative review (Littell et al., 2008). The exclusion criteria were structured hierarchically and assessed in order of importance.

Table 5.2: Inclusion and Exclusion Criteria – Effectiveness Review

Inclusion Criteria	*Exclusion Criteria*
1. RCTs (including cluster randomized trials) or quasi-experimental studies 2. Focused on interventions which have been funded by the SIB funding model 3. Includes primary data about the effect of SIB funding model and/or the intervention on service provision 4. Published in English 5. Retrievable	1. Wrong study design (i.e., the paper does not provide an empirical quantitative contribution) 2. Wrong intervention (i.e. not focused on intervention funded by the SIB model) 3. Does not include primary data about funding model or program level benefits (effectiveness) or harms 4. Not published in English 5. Not retrievable

In deciding whether a particular program qualifies as a SIB, the defining feature was the provision of upfront capital from external investors who earn a return if the intervention succeeds (Albertson et al., 2018). While SIBs represent a subset of PbR, broader literature relating to other types of outcomes-based commissioning was excluded. As SIBs are a relatively new funding mechanism and there is little published literature available, the inclusion criteria did not specify that articles need to be peer-reviewed and published in scientific literature. For practical reasons, only English-language documents were included. Although the first SIB was introduced in 2010, similar mechanisms may have been used before then, thus necessitating an earlier start date than 2010. January 1990 was deemed to be

appropriate, as this marks the start of the decade where experimentation in PbR and outcomes contracting became prevalent (Economy et al., 2020).

Search methods for the identification of studies

The standard search string was designed to retrieve a large number of studies related to SIBs. It was primarily organized around synonyms used to describe SIBs; in some databases, study-type terms were also used. Appendix A shows an example (SCOPUS) search string. The search terms used for the updated review incorporated suggestions made by Policy Advisory Group members as part of a broader systematic review project (Picker et al., 2021) and accounted for changing terminology surrounding SIBs. Additional steps (e.g. grey literature search) were also taken to locate studies that have yet to be indexed in the academic databases, as described below.

Databases search

Searching a diverse array of databases was essential to retrieve a broad range of published and unpublished literature. This included databases with peer-reviewed articles and databases containing dissertations and conference proceedings. From preliminary searches, 12 databases were found to be relevant. Beyond this number, there were diminishing returns for each additional database searched. Sensitivity maximizing searches were conducted in each of the following databases on 24th January 2021 (to update the 2018 review): ABI/INFORM Global; Applied Social Sciences Index and Abstracts (ASSIA); International Bibliography of the Social Sciences (IBSS); PAIS Index; PolicyFile Index; Proquest Dissertations and Theses Database; Proquest Social Science Database; PsycINFO; SCOPUS; Socical Services Abstracts; Web of Science Core Collection; and Worldwide Political Science.

Grey literature search

A considerable body of SIBs research has been conducted outside of academia. Thus, a grey literature search was an important component of the effectiveness review (Adams et al., 2016; Mahood Q et al., 2014). At the outset, reference lists of existing literature reviews were hand-searched to identify grey literature sources to search in more detail. I also cross-checked identified grey literature articles with an (unpublished) reference list compiled by experts from the Government Outcomes Lab (GO Lab) (Blavatnik School of Government, University of Oxford) and Ecorys (Picker et al., 2021). While this assisted with identifying grey literature sources and some additional studies, it is essential to recognize that this approach is "far from objective" as reference lists may selectively cite sources (Higgins et al., 2019). Appendix B summarizes the results of the grey literature search.

Hand searching, snowballing, contacting experts

A combination of electronic and hand searching is required to identify all relevant reports (Hopewell, 2007). This is integral as some reports are not included in electronic bibliographic databases, and others (even if included) may not be appropriately indexed (Dickersin, 1994). Consequently, the search strategy also included hand-searching each of the journals that have published articles on SIBs to date (Appendix C). Additionally, given that this is still an emerging field of research, authors who have published multiple articles on SIBs were contacted via email and asked if they could provide unpublished studies or general assistance identifying additional studies (Appendix D). In total, 54 authors were contacted in total. At the time of submission, the experts had suggested 53 additional articles, 14 of which had not already been identified.

Extraction of information

Study screening and selection

Studies identified from electronic searches and the additional searches were combined in Covidence (systematic review software), and duplicates were removed. I firstly screened abstracts and titles for potentially relevant studies. I excluded studies if, based on the title and abstract, they did not meet one or more of the hierarchical inclusion criteria. The remaining, potentially relevant reports were retrieved in full text and assessed for eligibility using the predetermined inclusion criteria.

Data extraction and risk of bias assessment of included quantitative studies

When extracting data for the effectiveness review, I used Cochrane's Effective Practice and Organisation of Care (EPOC) data collection form. All quantitative studies were critically appraised using a standardized quality assessment tool designed specifically for quantitative research (Appendix E). Specifically, I used a modified version of the Hamilton tool ("Quality Assessment Tool for Quantitative Studies" – Effective Public Health Practice Project 2007) (EPHPP, 1998). The tool consists of six questions relating to: 1) selection bias; 2) study design; 3) confounders; 4) blinding; 5) data collection; and 6) withdrawals and dropouts. Each of the questions can receive either an A (high), a B (medium) or a C (low) quality rating (Lorenc et al., 2014). Using this tool, each study was assigned to one of the following categories: 1) low risk of bias; 2) moderate risk of bias; or 3) high risk of bias. The risk of bias in each domain was presented separately for each study in Appendix H.

Quantitative synthesis

As there is a dearth of quantitative data in this field, it was not possible to carry out quantitative meta-analysis. The review identified a small number of experimental and quasi-

experimental studies containing relevant empirical data and significant heterogeneity of studies. Due to this limitation and the significant heterogeneity of interventions, evaluation designs, and outcome measures, I synthesized the effectiveness studies descriptively (i.e. narrative synthesis) rather than statistically (i.e. with meta-analysis). Narrative synthesis is an approach which has been used in many other studies with similar data limitations (e.g. Lorenc et al., 2014). Where available, effect sizes were analyzed as change scores (i.e differences between pre and post) and relative change scores (i.e differences between change scores in the intervention and control groups). Ordinal outcomes, such as Likert-type scaled, were standardized to 0-100. To summarize effect sizes across multiple distinct outcomes and/or studies, median change scores were calculated, thus providing an indicative measure of effect size for each study. The outcome data were synthesized separately for controlled and uncontrolled studies to ensure that any divergence between these two groups were identified. I summarized the information in a table which outlines basic intervention and study information, along with whether studies reported positive, null, adverse or mixed results (Appendix G). Figure 5.6 summarizes the effects. While a formal cross-study synthesis (using methods outlined by the EPPI Centre) was not undertaken, the findings of subsequent chapters consider how the data from this chapter support and/or conflict with findings identified through the other types of data. If the evidence base is significantly strengthened in future, this review should be updated, with statistical meta-analysis included in the synthesis.

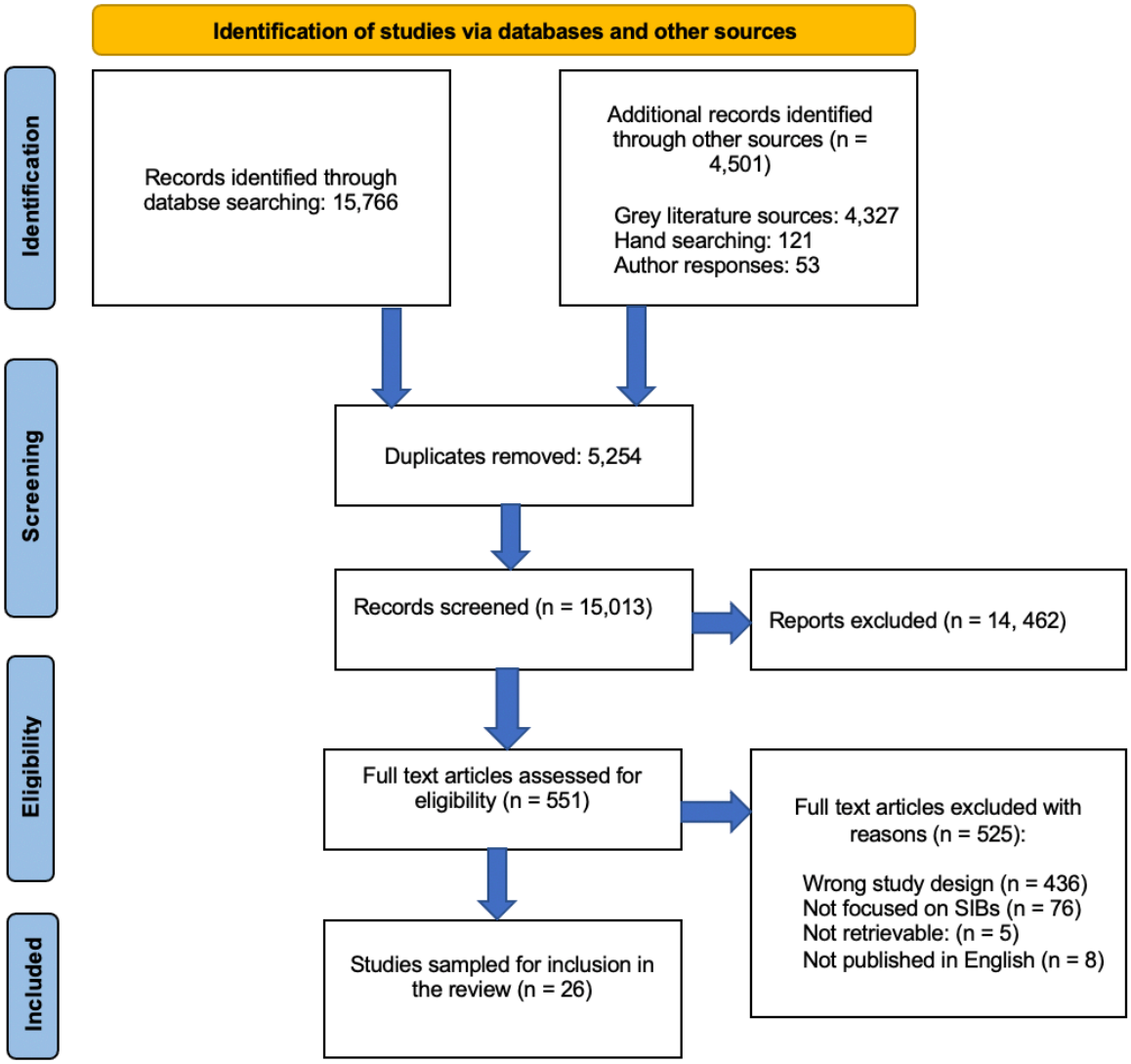
5.3 Results

5.3.1 Search Results

The initial search of electronic databases, grey literature sources and hand searches was conducted between 1st July and 20th July 2018. The search was updated between 24th and 25th January 2021 using revised search terms. Professional outreach was also conducted,

including contacting 53 academics who have published multiple SIBs articles to identify any additional or unpublished studies. Figure 5.1 shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram for the updated review. The searches of electronic databases yielded 15,766 citations (compared to 7,479 citations in the original review). I collected an additional 4,501 citations from other sources. After removing duplicates, 15,013 articles were retrieved for title and abstract review. I excluded 14,462 articles on title and abstract; 551 articles were sought for full-text review.

Figure 5.1: PRISMA Flow Chart of Study Screening



5.3.2 Excluded studies

For the revised review, I obtained the full text for 551 studies. Each of the studies would have been excluded if I had retained a non-SIB comparator requirement in the inclusion criteria (i.e. 'empty' review). Thus, I used the "best available evidence" on SIBs rather than restricting the analysis to the "best evidence" (Slavin, 1995; Ogilvie et al., 2005). Identifying and synthesizing the "best available evidence" is an approach advocated by many

academics, particularly when the evidence base is limited. (Slavin, 1995; Ogilvie et al., 2005).

5.3.3 Included studies

A total of 26 publications met the full inclusion criteria (compared to 16 in the original review) for the impact evaluation synthesis and were thus included in the analysis. The 26 publications represented 18 unique studies; I define a 'unique study' as one with a specific design and sample, which could have resulted in one or more publications (e.g. by reporting different outcomes or follow-up periods). Appendix G provides the complete list of publications, showing how they have been clustered together as unique studies in this chapter.

Focus of studies

As demonstrated in Appendix G, the evaluation objectives for the included studies did not align with the original review goal of identifying evidence of SIB (funding model) effectiveness. However, they did each aim to identify whether the SIB-funded intervention was effective at the program level (e.g. by reducing recidivism rates) and thus assisted with learning about possible instrumental motives. Ultimately, in practice and policy (e.g. those operating at higher levels of public administration), there is excitement about (and an over-emphasis on) the potential of the funding mechanism, with little attention paid to its effect on outcomes for recipients. However, in the literature, there is a strong focus on the impact of the SIB-funded interventions on recipient outcomes, with little analysis of the impact of the funding structure itself. This suggests that there is a lack of accountability between the two interest groups, who appear to be focusing on different research questions and agendas. The fact that each of the included studies focused on the effect of the specific intervention (e.g.,

recidivism-reduction programs) on service recipient outcomes, rather than the effect of the commissioning model on changes to service delivery outcomes (compared to other possible ways of commissioning services) highlights a significant gap in the evidence base. As there are still no studies that have quantitatively compared a SIB site to a non-SIB comparison site (for a similar intervention), it is unclear to what extent outcomes can be attributable to the funding model itself and uncertainty thus remains, more than 12 years on from the launch of the first SIB. Consequently, it is possible that many of the outcomes achieved in SIB-funded programs may have been achieved without the added layer and complexity of the SIB funding model.

Study locations

The largest number of unique studies stem from USA (n = 7), followed by Australia (n = 5) and the UK (n = 3). Several countries (India, Portugal, Kenya, and Uganda) only had one unique study. Given that the UK has launched the most SIBs (93 out of 276 launched globally), it is concerning that they have so few experimental or quasi-experimental evaluations (GO Lab, 2023a). Regardless, it is not surprising given that many UK SIBs have used a 'rate card' approach in which pre-defined outcomes are listed with an amount of money specified as payment for each successful outcome; validated administrative data are used to assess progress. Far fewer SIBs have been launched in the other countries represented in unique studies: USA (n = 28); Australia (n = 15); India (n = 4); Portugal (23); Kenya (n = 2); and Uganda (n = 1) (GO Lab, 2023a).

Study designs

Of the 18 unique studies, 44.44% (n = 8) used some level of randomization; in 2 studies, this included a cluster RCT design (Gallardo et al., 2021; ID Insight, 2018). The

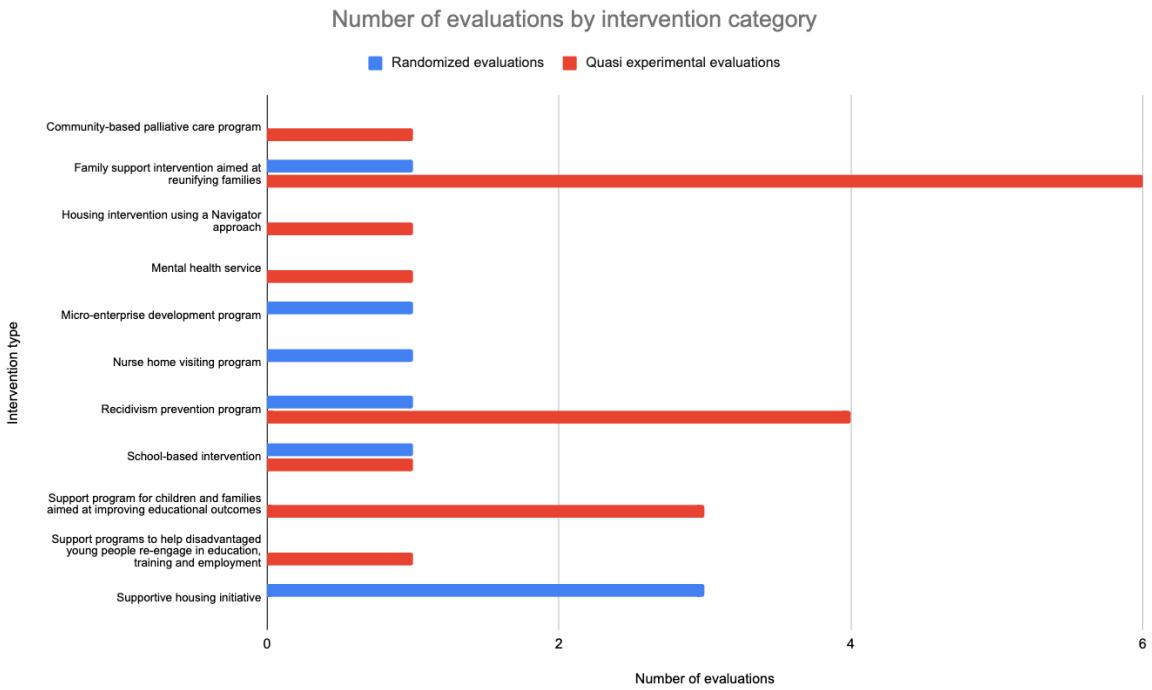
remaining studies used a non-randomized design with a comparison group. For five unique studies, this involved using propensity score matching (PSM), which is considered an approach that can yield a well-matched comparison group (Riker's Island SIB; Peterborough SIB; London Homelessness SIB; Innovation Fund Pilot; Chicago Child-Parent Center Pay for Success (PFS) Initiative)). Regardless, concerns were raised in at least one of the studies using PSM, about the comparability of the comparison group, due to data limitations (Salis et al., 2018). Five of the unique studies used a lower-rigour comparison (matched control group), in which it was unclear if the groups were comparable (Newpin Social Benefit Bond (SBB); Resilient Families SBB; Resolve SBB; Silver Chain SIB; Junior Code Academy)). Only two studies (Village Enterprise DIB and South Carolina Nurse-Family Partnership PFS Project) have published a study protocol (Carson, 2021; McConnell et al., 2020). Both of these studies were pre-specified and registered on the American Economic Association's registry for RCTs. The sample size in included studies ranged from 94 (Junior Code Academy) to approximately 12,000 (Educate Girls DIB). Only one included study was located through the database searches rather than the grey literature search, pointing to the lack of effectiveness studies indexed in academic databases (Parsons et al., 2016). This also raises possible issues with current efforts to use machine learning to make systematic review processes more efficient (van de Schoot et al., 2021). While it may be possible to speed up parts of the screening process, it appears that in some fields, the search process will remain laborious due to the lack of robust studies indexed in academic databases.

Interventions

The included studies evaluated 11 different intervention types (Figure 5.2). The most common intervention types were recidivism reduction programs, family reunification programs and supportive housing initiatives, each featuring in 3 unique studies. All other

intervention types were only featured in one unique study. The following intervention types lack evidence from a randomized study design: 1) community-based palliative care program; 2) navigator housing intervention; 3) mental health service; 4) support program for children and families aimed at improving educational outcomes; and 5) support programs to help disadvantaged young people re-engage in education, training and employment. The supportive housing initiative category contained the most randomized evaluations. The intervention duration was unclear for many of the interventions due to inadequate reporting. For those that did report the duration, the shortest timeframe was approximately three months; the Silver Chain program was delivered to individuals in their final three months of life (Paxon Group, 2020). For many programs, 2-3 years was a typical delivery timeframe, with the Chicago Child-Parent Center PFS Initiative lasting for longer as participants receive support from pre-school through to grade 3 (Gaylor et al., 2017; Gaylor et al., 2018; Gaylor et al., 2019). The intensity of the program was also rarely discussed. The most detailed reporting of this was for the Resilient Families program, where it was reported that the total average intensity increased after the Interim Report, from 11.9 to 12.4 interactions per month, with the average hours of contact per client each month increasing from 8.1 hours to 9.5 hours (Leahy et al., 2017). Most of this increase occurred for very high-risk families.

Figure 5.2: Evidence base for each intervention type[2]



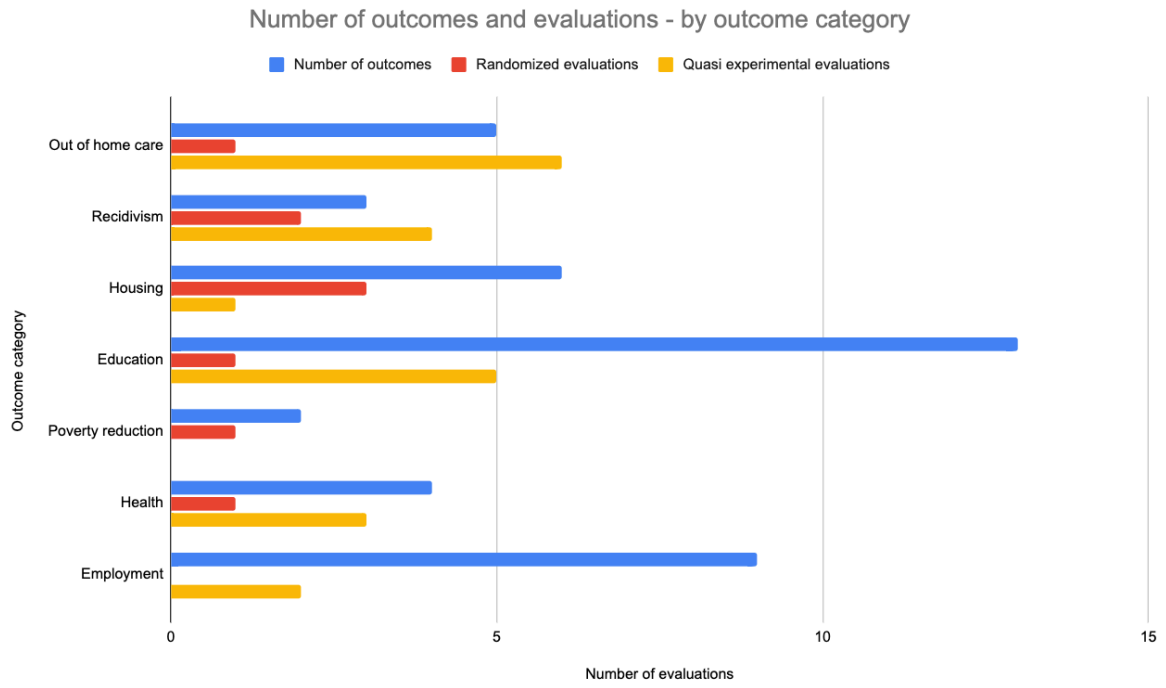
Outcomes

The outcome types are depicted in Figure 5.3. In total, 42 outcomes were identified from seven broad policy areas. Policy areas with a particularly high number of outcomes included education (n = 13); and employment (n = 9). Many evaluations featured similar outcomes that were operationalized differently, thus preventing the pooling of results. For example, the Riker's Island SIB and Denver SIB analyzed recidivism 'bed days'. However, this was measured in the 12 months post-release for Riker's Island (Parsons et al., 2016) and in the three years post-randomization for Denver (Cunningham et al., 2021).

[2] For this figure, each publication is counted separately (even if for the same project). In some instances, there are some overlaps that have not been captured (e.g. Denver is a housing program but it also contains a recidivism outcome).

Similarly, 'reduction in the consumption of health services' was a payable outcome in various SIBs but measured differently in each project. In the Resolve SIB, it referred solely to the percentage reduction in National Weighted Activity Units (NWAUs) (Wise et al., 2022). In Silver Chain, it considered admissions, Emergency Department (ED) presentations, bed days and NWAUs (Paxon Group, 2020); and in the London Homelessness SIB, it only referred to reductions in cohort Accident and Emergency admissions (DCLG, 2017). The lack of standardized measures severely limits the ability to pool results from different studies and should be addressed in future SIB-focused studies.

Figure 5.3: Number and types of outcomes[3]



[3] As with Figure 5.2, in this figure, each publication is counted separately (even if for the same project). In some instances, there are some overlaps that have not been captured (e.g. Denver is a housing program, but it also contains a recidivism outcome).

Timing of evidence

The timing of evidence is crucial to consider, given my focus on examining how changing levels of uncertainty have influenced the mix of efficiency and legitimacy drivers. In Figure 5.4, I present the chronology of evidence by showing the number of publications released per year, reporting a particular category of outcome. A small number of publications reported multiple outcomes (e.g. housing and reoffending); I counted these studies for each outcome published to allow for a more nuanced understanding of the outcomes reported. As depicted in Figure 5.4, there was a complete lack of quantitative evidence between 2010 – 2013. Evidence emerged from 2014 onwards, although the reports containing positive outcomes were typically outnumbered by those containing null, mixed, negative or unknown outcomes. This changed slightly from 2019 onwards, with more positive outcomes being reported, although by this point, the uptake of SIBs had already declined. Thus, the positive results may have had a limited impact, as the novelty of the innovation had likely already worn off, and attention had decreased.

Figure 5.4: Chronology of evidence

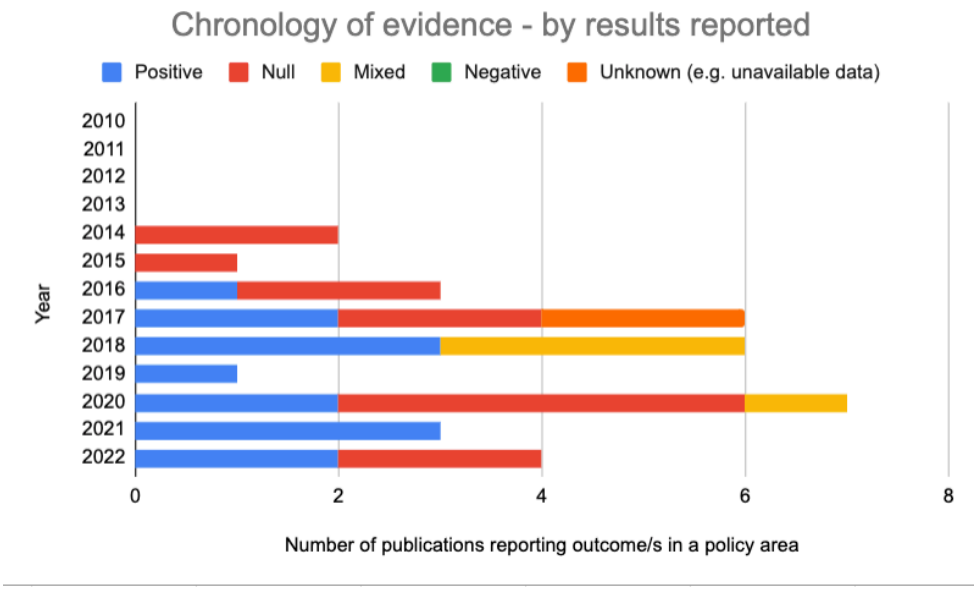
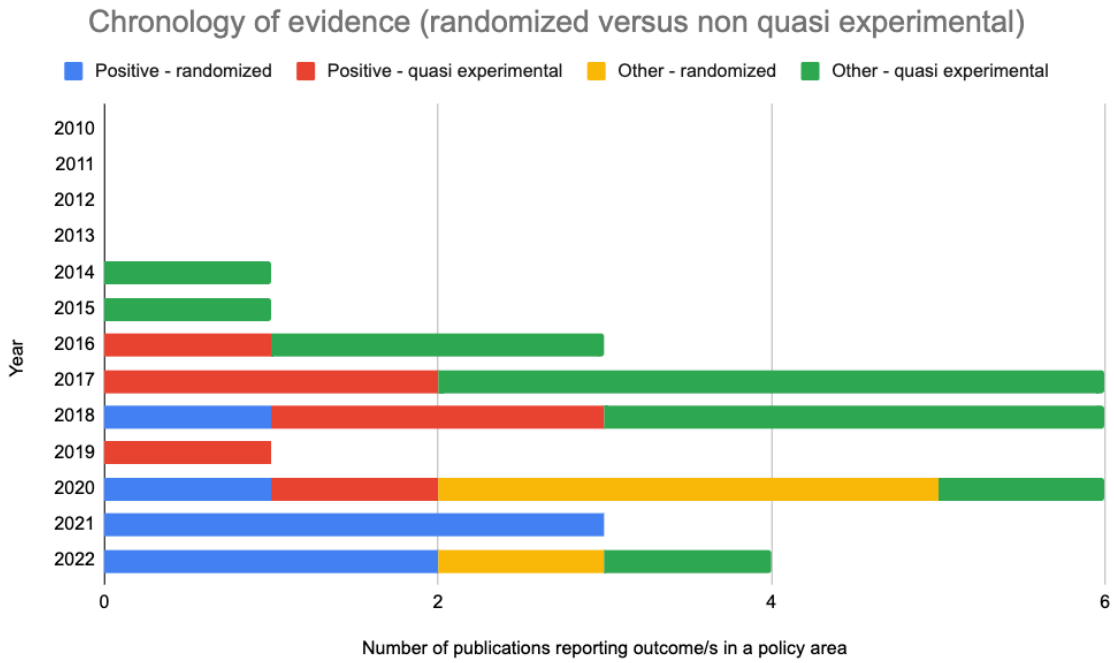


Figure 5.5 also shows the chronology of evidence, although it considers whether the relevant publication used a randomized or non-randomized design. This is an essential consideration, as weaker designs produce stronger results due to biases. As depicted in Figure 5.5, positive results were reported from studies with a randomized design in 2018 and between 2020 and 2022. Studies reporting a non-positive result (labelled as 'other' in the graph) more typically used a non-randomized design. Thus, Figure 5.5 demonstrates that while some positive (quantitative) findings emerged from 2016 onwards, these results were reported in studies with a weaker evaluation design. The evidence base only appeared to strengthen from 2018 onwards. However, this was eight years after the launch of the first SIB, and the lack of results leading up to this may have already been interpreted as a sign of ineffectiveness by SIB stakeholders, thus potentially contributing to the decline of SIBs.

Figure 5.5: Chronology of evidence (comparison of experimental and quasi-experimental results)



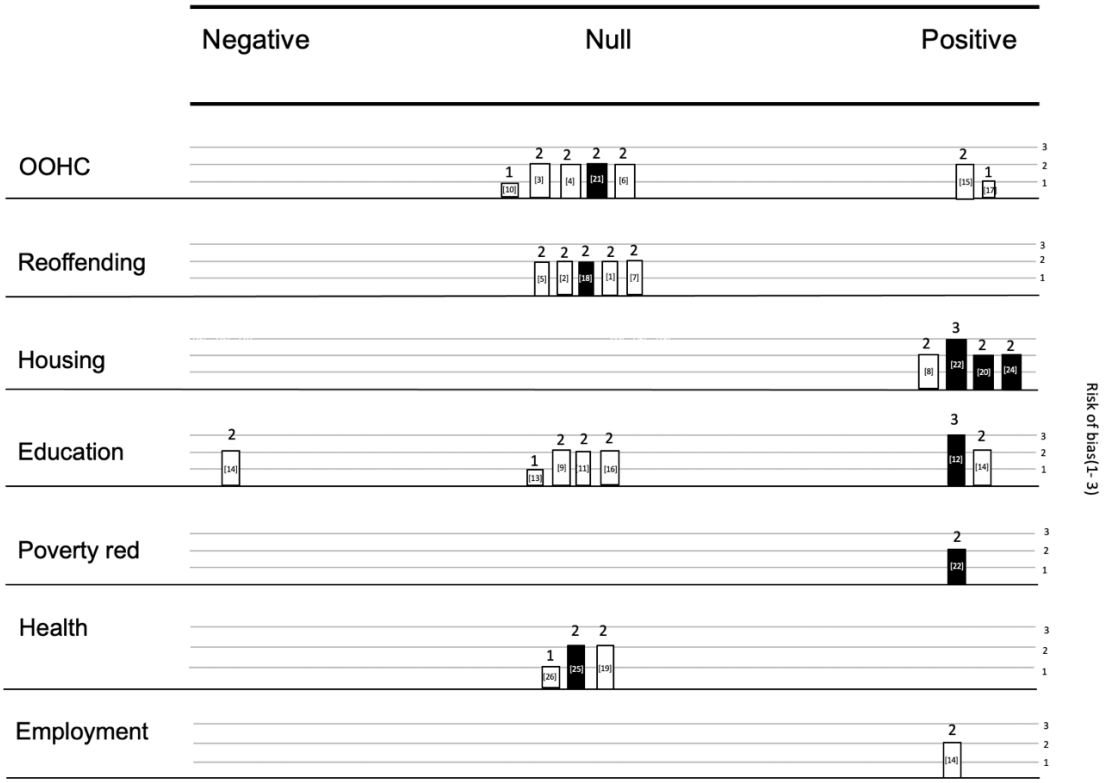
5.3.4 Effectiveness evidence

This review found evaluations of 18 SIB-funded projects aimed at enhancing a diverse array of social outcomes. Figure 5.6 provides a high-level summary of the

effectiveness evidence, grouped by outcome area, showing a clear clustering of 'null' results for many policy areas. Details of the specific interventions are included in Appendix G. Although there are some existing qualitative studies which suggest some differences in service delivery (e.g. greater flexibility in service delivery) between SIB and non-SIB comparator sites (e.g. Fraser et al., 2018), the scarce quantitative evidence makes it difficult to truly understand the causal connection between the SIB funding model and enhanced (or worsened) service delivery and outcomes. Some impact evaluations identified in the effectiveness review noted that despite some positive findings, it has not yet been possible to prove that the social investment and payment incentives are contributing to the results. This chapter thus demonstrates that existing evaluations have not disentangled the impact investment model from the intervention service, leading to a significant gap in the literature (and ongoing uncertainty) which should be prioritized in future research. In the subsections below, I discuss the results for each of the broad categories used to organize the available evidence. In narratively presenting the findings, I indicate whether the evidence for each policy area indicated positive, negative, mixed or null results.[4]

[4] Results were considered positive when all included studies showed statistically significant positive (i.e. favourable) results for the given outcome area. Mixed effects indicated that one or more studies found positive effects, and one or more studies found negative or null effects (or mixed effects within a single study). Null effects meant no statistically significant results were demonstrated for measured outcomes applicable to the given outcome area. Negative effects meant that the included studies demonstrated statistically significant negative (i.e. adverse) effects and no positive effects.

Figure 5.6: Harvest Plot Summarizing Effects[5]



[5]The harvest plot summarizes several aspects of the primary studies. Each block represents an individual study. The positioning of each block under one of the three headings reflects which of the three categories of results (positive, null, negative) is most supported by the findings of each study. The height of the bars represents the risk of bias in the study design (1 = high, 2 = mod, 3 = low), and the tone of the bars indicates whether the study used a randomized (black) or non-randomized (clear) evaluation method. The number within each bar is the citation number for the study (Appendix H).

Narrative Synthesis

Health

Three unique studies (reported on in three publications) evaluated the following SIB programs: Silver Chain; Resolve and the South Carolina Nurse-Family Partnership PFS Project (Paxon Group, 2020; Wise et al., 2022; McConnell et al., 2022). Two of these programs took place in Australia (Paxon Group, 2020; Wise et al., 2022), and one was conducted in USA (McConnell et al., 2022). The Silver Chain program delivers enhanced community-based palliative care 24/7 in the home; it features the shortest project duration, as it is primarily designed for people in their final few months of life (Paxon Group, 2020). In the Resolve program, beneficiaries were provided with a mental health service that blends psycho-social and clinical services (Wise et al., 2022). The South Carolina Nurse-Family Partnership PFS Project features an established model of nurse home visiting; regular visits begin prenatally and continue until children reach the age of two (McConnell et al., 2022). The London Homelessness SIB also featured a health outcome; however, it was not included in this category as it is primarily a housing intervention, and health impacts could not be determined due to data limitations (DCLG, 2017).

Several studies included an outcome related to a reduction in the consumption of health services, although this was operationalized differently. All other health outcomes were only included in a single SIB study (birth spacing; likelihood of having an adverse birth outcome; and the likelihood of experiencing an injury, abuse or neglect during early childhood) (McConnell et al., 2022). The studies included in this category involved varying levels of rigour, with one RCT (South Carolina Nurse-Family Partnership PFS Project; McConnell et al., 2022) and two studies that used a control group (Silver Chain and Resolve; Paxon Group, 2022 and Wise et al., 2022). A control group was originally included in the

Revolve program; however, in 2021, measurement was changed to a comparison with a beneficiary's experience in the year prior to their enrolment (i.e. pre/post program measure) (Wise et al., 2022). This change lowered the level of rigour, thus warranting a higher degree of caution in interpreting study results. For the Silver Chain and South Carolina Nurse-Family Partnership project, it was explicitly noted that the control group received usual care (Paxon Group, 2020; McConnell et al., 2022). It is unclear what the control group received in the Resolve program, for the time period in which the study design included a proper comparator (Wise et al., 2022).

Regardless of the study design, the results in this category were inconclusive, with two studies reporting null results (Silver Chain and South Carolina Nurse-Family Partnership PFS Project); and one study (Resolve) reporting mixed results. Thus, there is a lack of evidence of effectiveness, for SIB funded health interventions.

Education

Five publications focused on SIB-funded programs using an education intervention (ID Insight, 2018; Maze et al., 2018; Gaylor et al., 2017; Gaylor et al., 2018; Gaylor et al., 2019). Of the unique studies, one was conducted in India (Educate Girls), one in Portugal (Junior Code Academy), and one in the USA (Chicago Child-Parent Center PFS Initiative). Junior Code Academy was a school-based intervention, with coding classes taking place within school hours in the presence of the head teacher (Maze et al., 2018). In Educate Girls, volunteers were trained to deliver a child-centric curriculum one to five times a week to boys and girls in Grades 3-5; and households with unenrolled girls were identified and targeted with multiple interventions (ID Insight, 2018). The Chicago Child-Parent Center PFS Initiative provided half and full-day preschool education to three- and four-year-olds, along

with comprehensive family services designed to enhance learning at home (Gaylor et al., 2017; Gaylor et al., 2018; Gaylor et al., 2019). The Innovation Fund evaluation also included several education outcomes, although it is discussed in the employment sub-category (Salis et al., 2018).

There is some evidence of effectiveness in this sub-category, with the Educate Girls evaluation (analyzed with an RCT) finding positive results for learning gains and enrolment (ID Insight, 2018). In particular, the effects on learning gains were large and statistically significant, with students in the intervention group gaining, on average, 28% more than control group students (ID Insight, 2018). While the project was consistently on track to meet the enrolment target, progress on the learning target had lagged considerably. For instance, after two years (of the three-year SIB), Educate Girls had only reached half of the target; however, by the end of the project, the final target had been exceeded by 60% (ID Insight, 2018). SIB stakeholders have suggested that the DIB funding model assisted with this as *"the continuous feedback and analysis of data allowed us to identify learning lags in children and weaknesses in our own teams. This led us to redesign our curriculum from scratch and retrain our staff"* (GO Lab, 2019).

However, results from the other education interventions were mixed; and the evaluation designs lacked rigour. For instance, the Junior Code Academy SIB reported that intervention group students achieved a higher score compared to the control group in logical thinking and math learning performance and a lower score in Portuguese language learning. However, there were several flaws in both the study design and reporting. For instance, the treatment and control groups were not comparable in terms of baseline characteristics, as the evaluation noted that *"the treatment groups started from a lower point than the control*

groups" (Maze et al., 2018, p. 1). Further, the descriptive statistics lacked important information, at both baseline and post-intervention; at least one of the outcomes (Portuguese language learning) was only assessed at the end of the intervention. Likewise, the Chicago Child-Parent Center PFS Initiative evaluation claimed positive results (e.g. lower special education rates in the intervention group than the comparison group at all time points); however, it is challenging to support the inferences drawn due to severe reporting limitations (Gaylor et al., 2017; Gaylor et al., 2018; Gaylor et al., 2019). For instance, for the key outcomes, the reports only provided the relevant score for the comparison and control group and the percentage difference. Other vital statistics, such as the standard deviations, median, spread and shape of the distribution, were not provided. Thus, the effect of outliers and skew remains unknown; and it is unclear if the results are statistically significant. While not included in an education evaluation, one other study, which included education outcomes, found adverse effects regarding school attitudes and behaviours (Salis et al., 2018). Specifically, in the academic year after commencing the Innovation Fund pilot, participants from the 2012/13 cohort experienced a (statistically significant) increase of 5.3 percentage points in the proportion recorded as persistently absent; the number of individuals receiving one or more exclusions also increased by a similar figure (Salis et al., 2018).

Overall, the evidence in this category is mixed as one randomized study showed strong and statistically significant results (ID Insight, 2018), while the other evaluations used weaker designs and showed mixed results (with one instance of adverse results) (Maze et al., 2018; Gaylor et al., 2017; Gaylor et al., 2018; Gaylor et al., 2019; Salis et al., 2018).

Employment

Two unique studies used employment interventions, including the Village Enterprise DIB (Gallardo et al., 2021) and the Innovation Fund Pilot (Salis et al., 2018). The Village Enterprise DIB was conducted in India, while the Innovation Fund Pilot was conducted in the UK. The Village Enterprise DIB featured a micro-enterprise development program which provided the poorest members of the communities with an integrated package of cash transfers and training (Gallardo et al., 2021). The Innovation Fund included ten individual SIB projects; while the Fund evaluation discussed all of the projects to some extent, an individual program evaluation has yet to be conducted for each of the SIBs (Salis et al., 2018). The Innovation Fund projects were diverse, although they typically involved an employment-related component; for instance, the Nottingham Futures SIB provided beneficiaries with coaching and signposting to specialist services and an apprenticeship agency (Salis et al., 2018).

The evidence for employment interventions was mixed, with some positive results reported in an RCT and mixed results reported in the less-rigorous evaluation. Gallardo et al., (2021) found that the Village Enterprise program had a positive and statistically significant impact on monthly consumption and household net assets for participating households. For the consumption measure, treatment households consumed 9.9 USD (or 6.3%) more per month than the control group; and for the net assets measure, households in the treatment group had USD 40.5 (or 5.8%) more in net assets than those in the control group. Conversely, the Innovation Fund evaluation reported mixed results (Salis et al., 2018). For instance, one year after starting the Innovation Fund pilot, the proportion of participants who were in school or college or a paid job was lower than the comparator group; however, the IF pilot helped participants achieve a first National Qualifications Framework (NQF) level 1

qualification, predominantly one year after they started on the program (Salis et al., 2018). Several limitations were noted by the study authors, suggesting that the results should be interpreted with caution. For example, data limitations made it impossible to observe all personal characteristics and circumstances associated with a greater probability of being or becoming Not in Education, Training or Employment (NEET). Thus, the PSM approach could not produce a fully comparable comparison group (Salis et al., 2018). While the London Homelessness SIB intended to include some employment outcomes (e.g. sustained part-time employment; sustained full-time employment; sustained volunteering), it was not possible to assess these outcomes, as data on employment outcomes is not routinely recorded for all rough sleepers (DCLG, 2017).

Overall, the evidence on employment interventions was mixed, as one study used a robust design and reported positive and statistically significant results (Gallardo et al., 2021), while another used a lower rigour design and reported mixed findings (Salis et al., 2018).

Recidivism

Five publications (covering three unique studies) used a recidivism-prevention intervention (Parsons et al., 2016; Jolliffe & Hedderman 2014; Ministry of Justice, 2014; Anders & Dorsett, 2017; Ooi et al., 2020). As with many other categories, the level of rigour varied, and the evidence was mixed. One study was conducted in Australia (ON TRACC), one in the USA (Riker's Island), and one in the UK (Peterborough SIB). Each program was terminated prior to completion. While Peterborough was terminated due to a similar program being rolled out nationally (and thus, the control group being impacted), both the Riker's Island and ON TRACC SIBs were terminated due to null results. The interventions differed slightly. In the Peterborough SIB and the ON TRACC program, the combination of activities

provided for each offender was determined by service providers based on individual needs (Jolliffe & Hedderman, 2014; Ministry of Justice, 2014; Anders & Dorsett, 2017; Ooi et al., 2020). Conversely, in the Riker's Island SIB, adolescents received Moral Reconation Therapy (MRT), an evidence-based intervention focused on improved decision-making (Parsons et al., 2016). Thus, two of the three programs entailed a flexible approach with no explicit theory of change, making it difficult to evaluate whether particular elements of the ON TRACC or Peterborough SIB were effective.

The evidence base for SIB-funded recidivism reduction programs is also inconclusive. The program with the most robust evaluation design was the ON TRACC SIB. However, this study found no significant differences between the two groups on any of the reoffending measures examined (Ooi et al., 2020). Sub-group analysis found only one statistically significant result; non-Aboriginal offenders assigned to the ON TRACC service were five percentage points less likely to commit a personal, property, or serious drug offence (Ooi et al., 2020). However, this result should be interpreted cautiously as it is the only statistically significant reduction identified across several outcomes; and no underlying factors (e.g. specialty of service providers) were identified that could account for this.

Similarly, the Riker's Island evaluation (which used PSM) found that the SIB-funded program did not reduce recidivism rates for beneficiaries (Parsons et al., 2016). When external factors were taken into account, the recidivism rate for eligible 16- to 18-year-olds was statistically equivalent to the matched historical comparison cohort (Parsons et al., 2016). In the Peterborough SIB, positive results were reported in several evaluation reports, although the results were not statistically significant at the conventional 95% level (Jolliffe & Hedderman, 2014; Ministry of Justice, 2014; Anders & Dorsett, 2017). For instance, for

cohort two, the mean impact was a reduction of 0.141 reconviction events, which was only significant at the 90% level (Anders & Dorsett, 2017). Thus, it was not possible to reject the null hypothesis that there is no association between treatment and recidivism outcomes.

Overall, there is minimal evidence of effectiveness in SIB-funded recidivism reduction programs, potentially contributing to the fact that there are a limited number of SIB projects focused on this policy area (19 out of 276 projects launched), even though the first SIB featured a recidivism prevention program (GO Lab, 2023a).

Family reunification

Three unique studies (Newpin; Resilient Families; and The Cuyahoga Partnering for Family Success Program) focused on family reunification interventions, spanning seven publications (Wallace et al., 2016; Wallace et al., 2018; Leahy et al., 2015; Leahy et al., 2016; Leahy et al., 2017; Leahy et al., 2020; Crampton et al., 2020). Two of the SIB projects (Newpin and Resilient Families) were conducted in Australia, while one (The Cuyahoga Partnering for Family Success Program) was conducted in the USA. Each of the interventions involved a range of therapeutic and practical supports. The Cuyahoga program also featured housing support as program participants received faster housing placement than they would have otherwise (Crampton et al., 2020). The programs differed slightly in terms of intervention timing. For instance, in the Newpin program, parents already had at least one child aged six or under in out-of-home care (OOHC) (Wallace et al.,; Wallace et al., 2018). Conversely, in the Resilient Families program, the intervention was delivered when there were concerns about the safety and well-being of children, which, if not addressed, would likely result in the removal of the children (Leahy et al., 2015; Leahy et al., 2016; Leahy et al., 2017; Leahy et al., 2020). There were also some differences in the comparator and the

reporting of this. For instance, in the Cuyahoga program, the intervention group received conventional services plus the SIB intervention, while the comparator group only received conventional services. However, it is unclear precisely what 'conventional services' entailed (Crampton et al., 2020). In the Newpin SBB, the control group received 'business-as-usual' interventions which comprised a range of other restoration strategies and supports (Wallace et al., 2016; Wallace et al., 2018). In the Resilient Families SBB, the control group also received business-as-usual child protection services, although it was more explicitly noted that there are some similarities in the business-as-usual services received by the control group (Leahy et al., 2015; Leahy et al., 2016; Leahy et al., 2017; Leahy et al., 2020). Both the New South Wales Department of Family and Community Services (FACS) and The Benevolent Society (TBS) could refer control group and intervention group individuals to similar external supports that complement service delivery, thus potentially partly explaining the similarities in some program outcomes (e.g., a comparable average number of monthly interactions per client).

While more evidence exists for family reunification programs compared to other intervention types, the evidence base is still mixed. The Cuyahoga impact evaluation featured an RCT design; it found that while children in the treatment group had a lower rate of exiting custody compared to control group children, the difference was not statistically significant (Crampton et al., 2020). Specifically, there were no significant differences based on average days in Out of Home Placement (OHP), with slightly shorter average stays in the control group (802 versus 871 days) (Crampton et al., 2020). A noteworthy finding was that when the exit destination was considered, the results suggested that African-American caregivers in the treatment group were more likely to reunify with their child than African-American

caregivers in the control group (Crampton et al., 2020). This was an important finding as this population is disproportionately represented in child welfare contexts.

The non-randomized studies featured mixed evidence, with both positive and null effects reported. Leahy et al., (2016) found that the intervention children received more reports to the FACS Helpline than control children and had more Safety and Risk Assessments (SARAs) commenced, yet experienced slightly fewer statutory OOHC placements than control children. Leahy et al., (2017) found that while contact with the child protection system had reduced for intervention children over time, the reduction was only slightly less than that experienced by control group children who received business-as-usual services. In the final evaluation report, Leahy et al., (2020) found that the Resilient Families service effectively reduced the likelihood of OOHC placements, with 184 fewer intervention group children compared to control children entering care during the measurement period. This difference was statistically significant ($p<.05$) (Leahy et al., 2020). Likewise, in the Newpin SBB, the net restoration rate was much higher for the intervention versus the control group. For instance, in the second interim evaluation report, Wallace, Wise and Griffiths (2018) noted that between 1st July 2013 and 31st December 2017, 53% of intervention group children were successfully restored to their families, compared to 18% of children in the control group ($p<.01$). Further, they specified that *"with 95% confidence, the net restoration rate for children in the Intervention Group is between 29.3% and 40.0% higher than the Control Group"* (Wallace et al., p. iii). Although positive results were claimed, the study design was less rigorous than the RCT conducted in Cuyahoga (which found null results) (Crampton et al., 2020); and no effect sizes could be calculated for Newpin as critical statistical information (e.g. standard deviations) was not provided (Leahy et al., 2015; Leahy et al., 2016; Leahy et al., 2017; Leahy et al., 2020).

Overall, a greater volume of evidence exists for family restoration interventions compared to other intervention types. However the results vary considerably, potentially linked to the study design. For example, the most robust study found null results (Crampton et al., 2020), while the non-randomized studies reported some positive results, along with some null findings (Leahy et al., 2015; Leahy et al., 2016; Leahy et al., 2017; Leahy et al., 2020; Wallace et al., 2016; Wallace et al., 2018).

Housing

Housing intervention outcomes were reported in 4 publications, which discussed four unique studies (Cunningham et al., 2021; DCLG, 2017; Hunter et al., 2022; Raven et al., 2020). Within this category, there were two types of housing interventions: housing support using a Navigator approach (London Homelessness SIB); and supportive housing initiatives (Just in Reach; Denver SIB; and Santa Clara Welcome Home SIB). Three SIB projects took place in the USA (Just in Reach; Denver SIB; and Santa Clara Welcome Home SIB); and one in the UK (London Homelessness SIB). The London Homelessness SIB used a Navigator approach, in which service providers used a personalized approach and supported beneficiaries to access existing provision and achieve sustained housing (DCLG, 2017). Each of the remaining SIBs used a Permanent Supportive Housing (PSH) approach, involving subsidized housing with additional, voluntary support services (Cunningham et al., 2021; Hunter et al., 2022; Raven et al., 2020). The comparator varied considerably between the projects. In the Denver SIB, control group members received 'usual care services' in the community; the report did not specify exactly what this entailed (Cunningham et al., 2021). The details of the London Homelessness SIB comparator group are unclear (DCLG, 2017). In the Just in Reach project, control group members did not receive a permanent form of

housing through the PFS program or through any other housing program, thus avoiding the potential for overlapping services (Hunter et al., 2022).

Conversely, in Project Welcome Home, the control group received 'usual care', which could have involved similar services to the treatment group, thus potentially biasing the results (Raven et al., 2020). More than one-third of the control group received another form of subsidized housing (Raven et al., 2020). All control group participants received referrals to shelters and other homeless services and remained eligible to receive services provided for individuals experiencing homelessness in Santa Clara County, including any form of shelter and temporary or permanent housing (Raven et al., 2020).

There were multiple randomized studies in this category; thus, the rigour level was considerably stronger than for other intervention types. The randomized evaluations reported mixed outcomes within studies. For instance, Welcome Home reported that intervention group members had lower rates of psychiatric ED visits IRR 0.62; 95% CI [0.43, 0.91] and shelter days IRR 0.30; 95% CI [0.17, 0.53], and higher rates of ambulatory mental health services use IRR 1.84; 95% CI [1.43, 2.37] compared to control group members; but there was no difference in total ED or inpatient use, or jail (Raven et al., 2020). Similarly, in the Just In reach project, the program was associated with significant reductions in county jail use, probation, homelessness, and inpatient physical and mental health services; and significantly increased use of outpatient mental health services in comparison to the control group (Hunter et al., 2022). In the Denver SIB, the intervention group spent significantly more time in housing than the control group (560 more days of housing assistance across three years); however, null results were identified for other outcomes (Cunningham et al.,

2021). For instance, the differences between the two groups' uses of emergency medical services were not statistically significant.

Regardless of the mixed results within studies, housing outcomes were consistently positive in each of the randomized studies. In the non-randomized study (London Homelessness SIB), the intervention group had fewer episodes of rough sleeping compared to the control group (mean number of 9.2 rough sleeping contacts for the intervention group versus 13.9 for the control group) (p value 0.0145) (DCLG, 2017). The evaluation claimed that by extrapolating from the two-year results, around 3,900 rough sleeping episodes were avoided due to the intervention (DCLG, 2017). However, this should be interpreted cautiously; no effect sizes were presented, and they could not be calculated as key statistics, including the standard deviations, were not provided. Overall, there is some evidence of effective SIB-funded housing interventions, and the rigour underpinning the evidence in this category is more robust than in many other categories.

5.3.5 Risk of bias in included studies & effects of interventions

The risk of bias and effects of interventions were measured for the 18 unique studies (26 publications), using the process outlined in Appendix E, and scores are presented in Appendix H. In brief, the scores were calculated using a modified version of the Hamilton tool. It consists of six questions in total, relating to: 1) selection bias; 2) study design; 3) confounders; 4) blinding; 5) data collection; and 6) withdrawals and dropouts. While the tool is well-recognized and has been used in various studies, such as Thomson et al's (2013) review of housing interventions, it must be noted that some of the categories are not directly applicable and/or appropriate for SIB evaluations.

Before presenting the risk of bias scores, it is thus integral to consider the extent to which each of the categories in the selected tool are applicable to and appropriate for SIB evaluations. Overall, tools which prioritize and favour experimental and quasi-experimental designs (and aspects related to these designs) may not be fully approprite to apply to SIB projects, noting the inherent evaluation challenges that exist, which mean that many SIB focused studies will inevitably be labelled as low quality, despite potentially using the best available method, for the circumstancs. More specifically, some of the individual components of tools are not always applicable to (or easy to adhere to in) SIB projects. For example, blinding can be difficult, in practice, in many SIB projects which are delivered to deeply vulnerable individuals and in contexts where complete blinding may not be appropriate or feasible. Also, the high-profile nature of some SIB projects (e.g. Peterborough) has likely made blinding of participants difficult in practice. Yet, the tool penalises studies that do not blind both outcome assessors and participants (resulting in all included studies being assessed as high risk of bias, for the blinding category).

Regardless, many of the components of the Hamilton tool (and similar tools) are universally important evaluation aspects which should be achievable in and relevant to SIB projects. For example, while most studies ranked poorly on the ‘data collection’ and ‘withdrawals and dropouts’ categories, these categories involve consideration of important evaluation aspects which are relevant to SIB projects (e.g. attrition rate, piloting of data collection tool, validity of data collection tool etc). Thus, for these categories, the ‘high risk’ ratings reflect poor study design and/or reporting, rather than inappropriate or inapplicable categories. The ‘study design’ and ‘confounders’ categories were also largely relevant (although not necessarily easy to achieve in a SIB project), as they considered factors such as the inclusion of a control group, random allocation, the extent to which groups are matched

on key variables etc. Likewise, the 'selection bias' category is an important component as it considers the extent to which the study sample likely represents the population from the relevant target area (i.e. the target population for whom the intervention is intended). Examples of possible types of biases related to this are referral bias filter (e.g. asthma patients from a teaching hospital may differ to asthma patients from a general hospital) and volunteer bias (e.g. people who volunteer to participate may have different outcomes from non-volunteers) . SIBs are often delivered to a small cohort of highly disadvantaged and vulnerable individuals; how participants are recruited into a project (and thus the risk of selection bias) differs by SIB project. The poor ranking received by each of the studies for this factor was thus not always reflective of the representativeness part of the criteria (as this differs considerably by project), but rather the fact that the baseline response rate (which also influences the score) was not reported in the included studies (i.e. the generalizability was often unclear; and/or the response rate at baseline was not reported or was unclear).

While the risk of bias tool had several shortcomings when applied to the SIBs context, it is worth noting that this was the case regardless of which tool was selected. There are well-documented issues with prescriptively using checklists or scales, which often prioritize particular types of evidence (e.g. randomized designs) and overlook the nuances of particular contexts (Higgins et al., 2019). While I used an adapted version of the Hamilton tool, it will be important for future SIB-focused studies to consider how risk of bias tools (and other evaluation tools) should be tailored to better suit the SIBs context.

Risk of bias was high in four publications (Leahy et al., 2017; Leahy et al., 2020; Maze et al., 2018; Wise et al., 2022); moderate in 21 publications (Anders & Dorsett, 2017; Crampton et al., 2020; Cunningham et al., 2021; DCLG, 2017; Gallardo et al., 2021; Gaylor

et al., 2017; Gaylor et al., 2018; Gaylor et al., 2019; Hunter et al., 2022; Jolliffe & Hedderman, 2014; Leahy et al., 2015; Leahy et al., 2016; McConnell et al., 2022; Ministry of Justice, 2014; Ooi et al., 2020; Parsons et al., 2016; Paxon Group, 2020; Raven et al., 2020; Salis et al., 2018; Wallace et al., 2016; and Wallace et al., 2018); and low in only one publication (ID Insight 2018). There were severe reporting limitations in many studies. For instance, only one study explicitly reported on blinding, yet it noted that neither staff nor participants were blinded to the intervention group (McConnell et al., 2022). Similarly, only one study discussed the issue of attrition in considerable detail, although it noted that it was not relevant as no sample was taken at baseline; thus, the researchers instead tracked and investigated the related concept of nonresponse at endline (Gallardo et al., 2021). Each of the publications had different research questions, objectives and outcomes compared to this effectiveness review. They considered the efficacy of the intervention, although to varying levels of rigour; however, none of the study objectives were nested in a broader question about the effect of the SIB funding model. Despite this, I summarized the key information, as they represent the best available evidence on the topic.

Of the 26 included publications, only seven provided sufficient information for effect sizes to be calculated (ID Insight, 2018; Jolliffe & Hedderman, 2014; Ministry of Justice, 2014; Anders & Dorsett, 2017; Leahy et al., 2016; Ooi et al., 2020; and Paxon Group, 2020). For the Educate Girls SIB, the effect size was only provided for the learning gains outcome; no effect size was provided (or possible to calculate) as the Working Group decided against estimating enrolments in control villages. Similarly, Leahy et al., (2016) provided an effect size for some outcomes (number of helpline reports; and SARAs commenced) but not for others (entries into OOHC). The remaining studies provided some information about the effect on recipient outcomes, although they did not report critical statistical information. No

studies provided data on the effect of the SIB funding mechanism on changes to service delivery outcomes.

In the absence of efficacy certainty, the intervention is likely still being propelled forward by legitimacy drivers; whereas with stronger evidence, it may have been shown to be ineffective earlier and abandoned due to disillusionment. With uncertainty still surrounding the efficacy of the funding model, it seems likely that SIBs represent an intervention for which we may never know, with certainty, whether it 'works' or not. In relation to my key hypothesis (i.e. that levels of legitimacy have affected the spread of SIBs), I argue that this review has demonstrated a lack of effectiveness for the funding model, making it increasingly unlikely, over time, that stakeholders would have pursued the idea for instrumental reasons. While a lack of evidence (and some negative results) may intuitively also disprove institutional motives (as this could cause SIBs to go from being legitimate to illegitimate), the evidence has largely been reframed and/or overlooked by SIB proponents (as demonstrated in Chapters 7 and 8). Further, I argue that the lack of evidence, highlighted by this review, has paved the way for rhetoric and persuasiveness to drive the diffusion of the idea.

5.4 Discussion

This effectiveness review demonstrates how the landscape of empirical knowledge surrounding SIBs has developed since the first SIB was launched in 2010. It shows how uncertainty has changed over a 12-year period (between 2010 – 2022). Diminishing levels of uncertainty do not always imply positive results, as there could be increasing levels of certainty that the intervention does not work as intended, over time. Equally, a lack of evidence (which contributes to ongoing uncertainty) does not necessarily mean that an idea does not work; the complexity of the intervention or policy area may make it challenging to

produce such evidence. While an ambiguous signal, I argue that in the context of SIBs, the lack of evidence potentially signals that the reform has failed to live up to expectations. Consequently, while early adopters may have pursued the reform for instrumental reasons, I argue that it has become increasingly likely, over time, that governments have pursued the funding model to secure legitimacy among valued stakeholders such as voters (e.g. by pursuing a novel idea).

Although the evidence base is still lacking more than 12 years on from the first SIB, there has been a recent resurgence of political interest in the funding model. Stanworth and Hickman (2022, p. 1) released a report that *"provides an assessment of the total public value so far created by Social Outcomes Contracts in the UK".* The report received significant political attention, for instance, with Boris Johnson enthusiastically promoting the reported ten-fold return on investment for the government. The report notes that the cost of outcomes achieved in the 72 included Social Outcomes Contracts (SOCs) (early-stage projects and those with data availability issues excluded) was £139 million and that £1.418 billion was generated in value, representing a benefit-cost ratio of 10.2. While this claim has been widely promoted by Big Society Capital and various politicians, the report has a number of limitations, thus warranting caution and scepticism. Firstly, it was commissioned by Big Society Capital, who have a vested interest in the success of the SIB model, having invested in 38 of the existing SIBs. ATQ consultants conducted the evaluation; they have also promoted themselves as SIB specialists and won multiple SIB contracts. While the findings of Chapter 8 will demonstrate the importance of taking into account COI, these conflicts were not sufficiently addressed in the report. Secondly, the report does not account for the fact that some outcomes may not have been attributable to the SIB-funded intervention (or the SIB funding model itself). The issue of causal inference is critical, given that *"expert opinions,*

pooled judgements, brilliant intuitions, and shrewd hunches are frequently misleading" (Stanley, 1957, pp. 195-201; Oakley, 1998). However, in this study, the authors explicitly decided not to take into account 'deadweight' in their overall value calculations. They justified this by claiming that the included studies have already accounted for 'deadweight'; and in some contracts, "*a persuasive case can be made for there being very little deadweight because the outcome would be very unlikely to occur without the specific intervention funded by the SOC"* (Stanworth & Hickman, 2022, p. 12). However, when analyzing the evaluation methods used in existing SIBs, it becomes clear that this is not an accurate assumption. Third, the report fails to acknowledge the well-documented evaluation challenges and the lack of positive results in the evidence base; as demonstrated in this chapter, for SIBs that have been evaluated with more robust methods (e.g. RCT), there is a lack of evidence of effectiveness.

Additionally, for many SIBs and Outcome Funds, the evaluators have been unable to proceed with an impact evaluation (e.g. Ronicle & Smith, 2020). Fourth, the report noted that *"we collected data on these projects from intermediaries and Investment Fund Managers (IFMs) who manage projects, supplemented by data from published evaluations of some programmes and projects"* (Stanworth & Hickman, 2022, p. 1). However, the data they collected is not provided in full. Fifth, while the evaluators assign "high confidence", "medium confidence", and "low confidence" to the differing amounts of value projections, these categorizations are not linked to factors such as the inclusion of a comparator. For example, in one instance, "high confidence" was assigned as *"a contract that enables children in residential care to move to foster care is certain to create value because foster care costs are always lower than residential costs and the value occurs as soon as the move is made, and for the period validated by the contract"* (Stanworth & Hickman, 2022, p. 9).

Finally, the cost calculations do not appear to include the well-documented added transaction costs that SIBs inherently entail (e.g. consultancy fees, legal fees etc).

To better appreciate the limitations associated with reports such as the one discussed above, it is vital to understand the full landscape of SIB evaluations. Importantly, this study has provided an exhaustive review of the available quantitative evidence. As no studies were identified that quantitatively compare SIBs to appropriate (non-SIB) comparators, the original goal of understanding how the SIB funding model affects service provision remains largely unanswered. Ideally, future evaluations should consider the effectiveness of the SIB funding mechanism, as well as the effectiveness of individual interventions. Without efforts to evaluate the overall funding approach, it remains unclear whether SIBs provide a beneficial way of organizing public services and whether the funding structure is contributing to enhanced service provision and outcomes for beneficiaries. In the absence of evidence, when a high level of uncertainty still exists, stakeholders such as policymakers may be *"willing to place an enormous amount of trust in the collective knowledge of the crowd"* (Cialdini, 1993, pp. 131-132; Rao et al., 2001). Rather than basing an adoption decision on concrete results and data, it may simply be presumed that *"if a lot of people are doing the same thing, they must know something we don't"* (Cialdini, 1993, pp. 131-132; Rao et al., 2001). This can be problematic and lead to errors because *"quite frequently the crowd is mistaken because they are not acting on the basis of any superior information but are reacting, themselves, to the principle of social proof"* (Cialdini, 1993, pp. 131-132; Rao et al., 2001).

An essential component of this thesis is to understand what happens to efficiency and/or legitimacy motivations when the level of uncertainty surrounding the effectiveness of

an innovation changes. The extent to which the 'true value' of an idea remains uncertain after adoption (and what the true value is) is a crucial factor in the extent to which an idea becomes institutionalized, with uncertainty assisting with institutionalization, by preventing an idea from being quickly abandoned due to disappointment (Rao et al., 2001). In comparing the original effectiveness review (conducted in 2018) to the updated one (conducted in 2021), it is worth noting that although the number of retrieved studies increased significantly (as broader search terms were used in the updated review to account for changing terminology), the number of qualifying studies only increased slightly, and the results remained very mixed. Thus, diminishing levels of uncertainty (about whether SIBs are technically beneficial or not) are not coming from evaluations definitively providing evidence on the SIB effect; instead, it is the lack of results more than 12 years on from the initial launch which suggests that the funding model may not be achieving its policy ambitions. I argue that while stakeholders may have adopted SIBs for efficiency reasons at the beginning of the diffusion cycle, it is less likely that they would now be pursuing SIBs for technical benefits, as there is a severe lack of evidence pointing to the technical gains.

In many policy domains and situations, stakeholders may emphasize that we need more and/or better evidence and that a lack of evidence does not necessarily mean that an intervention is not effective. In the SIBs context, however, if legitimacy motives are still contributing to the spread of the idea, simply striving for 'more' and/or 'better' evidence may not be the solution. If stakeholders are adopting the innovation for reasons other than efficiency motivations, they may ignore or reframe the evidence and/or shift the innovation to a new context. This is echoed in the cautionary literature on SIBs; for instance, Saltman (2017, p. 14) notes that *"The solution to the problem of Pay for Success is not more efficacy studies, debates about the metrics, or call for further research and experimentation. Such*

approaches affirm rather than question the myths that Pay for Success is premised on. Rather, Pay for Success needs to be recognized as a project that represents the financial interests of a small number of ruling class people to hold public service provision hostage to private sector profit taking".

More broadly, outside of the SIBs context, there are many examples of evidence having little influence on policymaking, despite the best efforts of policymakers. For instance, in a qualitative study conducted by Smith (2013a), interviewees noted that early years policies had been pursued for various reasons completely unrelated to evidence, such as the personal belief held by several senior ministers, that children should be a policy focus; and a perception that policies focused on children were less likely to be criticized by voters or journalists. Thus, *"while some ideas demonstrate a remarkable degree of persistence within policy, others appear to falter or splinter as they move from research into policy; variations which do not appear to relate to the strength of the underlying evidence"* (Smith 2013a, p. 109). Specifically, in the context of SIBs, the shift from pursuing SIBs in developed countries to launching DIBs in developing countries could be a deliberate attempt to shift attention from the lack of evidence and to promote the idea to a new, enthusiastic group of stakeholders.

A lack of evidence provides an ambiguous signal regarding instrumental versus institutional motives, particularly when the emerging literature on complex interventions has been considered. Regardless, it was still important to conduct a systematic review at the outset to understand precisely when positive and negative evaluation findings have emerged. This was crucial for identifying the points at which legitimacy and efficiency drivers are likely to have shifted. The review found that the early years of the diffusion cycle (2010-

2013) were characterized by a complete lack of quantitative evidence. From 2014-2017, studies emerged using a non-randomized design and predominantly reported null results. From 2018 until now, some positive results have emerged (using randomized designs), although counterintuitively, this period coincides with a decline in the usage of SIBs (GO Lab, 2023a). Thus, as SIBs usage continued to grow in the absence of evidence and has declined alongside positive results, there is no clear link between instrumental motivations and the adoption of SIBs.

In discussing the evidence base surrounding SIBs, it is integral to note that SIBs are implemented in complex adaptive systems, where interventions are likely to have *"diverse, far reaching and non-linear effects"* (Shiell et al., 2008, p. 1283). Attention has been drawn to such issues in recent public health debates and literature relating to complex systems models of evidence (Rutter et al., 2017). Within the field of public health, the evidence underpinning responses to many contemporary public health challenges has been generated using tools and methods that were developed to answer questions about the effectiveness of clinical interventions (Shiell et al., 2008). Therefore, the tools and methods are based on linear models of cause and effect. Evaluation of outcomes often involves measuring changes at an individual level and summing those to capture the 'social' impact. The properties of a complex system cannot simply be predicted from the elements within it, as they are "*more than just the sum of its parts"* (Rutter et al., 2017, p. 1). Consequently, a complex systems model of public health: 1) conceptualizes health inequities as outcomes of 'a multitude of interdependent elements within a connected whole'; 2) recognizes that these elements affect each other in subtle, often non-linear ways; and 3) uses a broad spectrum of methods to evaluate interventions in recognition of this complexity (Rutter et al., 2017, p. 1).

Within the context of SIBs, it is challenging to use a simplistic model of cause and effect, as the 'intervention' is not necessarily a single event resulting in a clearly discernible outcome. SIBs involve a complex mix of stakeholders, commissioning processes, social interactions, conditions, and environments. Thus, it is crucial to consider the ways in which the processes and outcomes drive change, at all points within the system. While evidence is often reframed and/or overlooked in the context of SIBs, further exploring how to effectively evaluate systems level change would still be worthwhile in this emerging field. It will be integral for producing evaluations that better account for the complexity and systems level interactions and could contribute to better understanding the extent to which instrumental motivations have been driving SIB adoption.

In interpreting the results of this study, both the strengths and limitations of the review must be considered. In terms of strengths, this systematic review was conducted in a rigorous and systematic manner, following PRISMA guidelines (Moher et al., 2009). I used a comprehensive search strategy to ensure that the search was exhaustive. This included searching 12 academic databases and various grey literature sources in both an original and updated review. Unlike many other studies (e.g. Albertson et al., 2018; Tan et al., 2015; Fraser et al., 2016; Fraser et al., 2018), the review was not limited to a particular jurisdiction or policy area. Much of the existing SIB literature has been produced by parties with a strong interest in the success of SIBs (e.g. intermediaries, consulting firms etc). Thus, studies often present strong arguments in favour of SIBs without using robust, empirical data (Wilson, 2014). Consequently, another strength is that this study was undertaken from a more independent perspective and systematically.

Regardless of the above strengths, there are also various limitations. As highlighted previously, most of the included studies had a different objective to this review. Evaluations providing some detail about the 'SIB effect' have primarily used interview data and were thus overlooked for this particular review even though relevant (e.g. Roberts & Cameron, 2015). For pragmatic reasons, studies were not double-screened, and the search was limited to English-language documents. However, there are several potentially relevant non-English publications, as SIBs have been launched in 38 countries. Further, publication bias may have affected the results. For example, many stakeholders have a strong imperative for SIBs to 'work'. Thus, evaluations may not be published if they do not show that SIBs are an effective funding model; some unpublished evaluations may have quantitative data of relevance to the research question. Due to the heterogeneity of studies, it was not possible to conduct a meta-analysis or to narratively compare data on the primary outcome of interest (i.e. effect of SIB model on service delivery) from different evaluations. I considered comparing the results of the best available experimental and quasi-experimental studies to the results of studies focused on other types of PbR (to create a comparison group). However, this was infeasible, due to a lack of comparable studies using different funding models. Similar data challenges have been noted in existing studies. For instance, Fraser et al., (2018) aimed to quantitatively assess the 'SIB effect' using SIB and non-SIB comparators (although only for a small number of UK SIBs). However, they could not identify comparison sites with the following minimum requirements: 1) appropriate counterfactual; 2) sufficient sample size; and 3) availability of relevant data. Thus, they did not conduct the quantitative part of the evaluation. Noting the inherent limitations in the available quantitative evidence, I will build on the results by incorporating qualitative literature in the following chapter.

5.5 Conclusion

In this chapter, I have described the findings of a systematic review of SIBs-focused effectiveness studies, aimed at understanding how the landscape of SIBs evidence has changed over time. The findings highlight that uncertainty remains regarding our understanding of the 'SIB effect'. While some evidence of program-level outcomes has emerged, it is still unclear how (if at all) the SIB funding mechanism affects service delivery and outcomes. These findings relate to the overall theme of uncertainty as they suggest that, in line with the existing (broader) literature, the difficulties inherent in measuring SIBs and the lack of evidence may have initially assisted with the idea continuing to spread rather than being abandoned. In this context, over time, a lack of evidence could be interpreted as providing greater certainty that the idea does not work. A dearth of evidence does not necessarily mean that there is an absence of true intervention effects, as it could be reflective of the complexity of the intervention or the issue, making it difficult to create such evidence. Yet, evidence is such a powerful rhetorical device when policies and public spending are being decided that absent other drivers, this can (and does) tend to result in diminishing levels of enthusiasm for popular initiatives. For example, delivery units (i.e. small teams that gather and analyze performance data on a single issue and report directly to a head of state) quickly became *"a multi-million-pound industry exported around the globe"*; however, in many countries the reform did not live up to the expectations of political leaders, leading to units being subsequently (and often silently) axed (Institute for Government 2017). I argue that in the context of SIBs, the lack of evidence (demonstrated in this chapter) has not yet resulted in a similar level of disillusionment and abandonment, due to the strong and unwavering institutional motives, which have contributed to evidence being de-prioritized and often ignored.

Regardless of the lack of evidence, SIBs usage did not decline markedly until 2019, almost 10 years after the introduction of the first SIB. The more recent decline suggests that legitimacy benefits alone can only propel an idea forward for so long; and that disillusionment may now be setting in, in the absence of robust evidence. As there are many challenges associated with measuring the SIB effect, it is difficult to disentangle whether the lack of evidence points to certainty of the model not working; or whether it is simply not possible to demonstrate that it works. The evaluation challenges have likely also assisted proponents of the idea, with maintaining enthusiasm for the SIB funding model.

While this chapter has provided some initial evidence of institutional drivers likely influencing the spread of the idea, given the absence of evidence of effectiveness, it will be essential to build on this, with other forms of data. Analyzing as many different data sources as possible is in line with the 'consilience' approach used in this thesis. It is particularly important in fields such as this, where there are severe deficiencies in the evidence base (Hood and Dixon 2010). Thus, in the next chapter, I conduct a systematic review of qualitative evidence regarding broader (non-efficiency) reasons for adoption.

Chapter 6: Qualitative Systematic Review

6.1 Introduction

In the previous chapter, I presented the findings of a quantitative synthesis of the effectiveness of SIBs. I discussed what this reveals about how levels of uncertainty have changed over time regarding the effectiveness of the innovation. I noted the importance of synthesizing the available quantitative evidence, as SIBs have been adopted in several public sectors since 2010, despite the absence of evidence on their effectiveness (Sinclair et al., 2019; Williams, 2019). I identified that existing evaluations tend to focus on evaluating the effect of a SIB-funded intervention (e.g. homelessness program) on specific outcomes relevant to the intervention (e.g. homelessness rates) (e.g. DCLG, 2017). There are currently no studies that quantitatively compare a SIB to an appropriate non-SIB comparator to understand the effect of the funding model itself. Consequently, I identified that program-level outcomes have been mixed (positive, null and adverse effects reported), and uncertainty remains regarding the impact of the funding model. Thus, while there are many 'promises' associated with SIBs, including better outcomes, there is a severe lack of quantitative evidence to support this.

In the absence of evidence of effectiveness, I argue that it is less likely (particularly over time) that stakeholders would be pursuing SIBs solely for instrumental reasons. Thus, it is crucial to consider broader possible reasons for adoption, which are not based on instrumental factors. While there is a need for more quantitative data, and the evidence base is often criticized, many rich qualitative studies exist. There is yet to be an exhaustive qualitative review and accompanying thematic synthesis of the qualitative evidence. Before we have an improved quantitative evidence base and/or concrete empirical data on the 'SIB effect' (which could support or disprove instrumental reasons for SIB adoption), a critical gap

to fill is to consider the emerging theory on the motivations for SIB adoption. While the quantitative evidence is underdeveloped, it is possible to look to the qualitative evidence base to see what is being said and/or implied about why SIBs are being pursued.

In this chapter, I identify, appraise and synthesize qualitative studies exploring SIB practitioners' views and experiences regarding the motivations for adopting SIBs. Examining the lived experiences of SIB practitioners allowed for the identification of factors which may have influenced SIB adoption in the absence of evidence of effectiveness. I systematically identify the relevant qualitative literature and analyze the studies using thematic synthesis (Thomas & Harden, 2008). I identify that, while there is a dearth of quantitative evidence on SIBs, and thus, the effect of the funding model is unclear, several of the included qualitative studies cited instrumental reasons, such as improving efficiency, as a main motivation for pursuing SIBs (e.g. Hughes & Scherer, 2014; Ronicle et al., 2014). In particular, many of the instrumental themes identified centred around how SIBs are worthy of pursuing as they can improve outcomes, achieve better value for money and enhance evaluation/accountability. Regardless, juxtaposing the findings of the effectiveness review (chapter 5) with this thematic synthesis demonstrates that there is a high level of decoupling (i.e. a gap between policy and practice) regarding the instrumental factors (Bromley & Powell, 2012). This is an important finding given that once an idea is decoupled from its evidence base, it is *"more easily open to differing interpretations and uses by various actors*" (Blyth, 2001; Smith 2013a, p. 77).

Regarding the instrumental motives identified in the qualitative studies, there is a lack of robust quantitative evidence for all of the identified themes; and in some instances, there is evidence of the opposite eventuating in practice (e.g. perverse incentives and/or harm rather than improved outcomes) (Fraser et al., 2016). In terms of evidence of institutional drivers,

the thematic synthesis uncovered some clear themes relating to the well-documented legitimacy pillars, thus suggesting that institutional factors have played a vital role in the growth of SIBs (Scott, 2013; Suchman, 1995). There was a noticeable difference in the extent to which different pillars of legitimacy were discussed, suggesting that some (normative and pragmatic) pillars have played a more significant role in driving SIBs adoption than others. However, when analyzing the results over time, no strong relationship was found between the factors identified and the timing of publication. As this thesis uses a consilience approach, which combines multiple sources of evidence, this chapter is followed by an analysis of social media data.

6.2 Overview of methods

This qualitative review aims to understand broader motives for adopting SIBs by asking: *"what are the perceptions of SIB practitioners regarding their motivations for adopting SIBs?"* (Table 6.1). The research question was constructed using the 'Sample, Phenomenon of Interest, Design, Evaluation, and Research type' (SPIDER) format, which focuses less on interventions and more on study design and samples (Cooke et al., 2012).

Table 6.1: SPIDER Research Question

Element	
Sample Phenomenon of interest	SIB practitioners who have been (or are currently) involved in delivering a social intervention that has been funded using a SIB structure
Phenomenon of interest	SIB-funded projects (i.e. projects that have involved the provision of upfront capital from private investors)
Design	Interviews, surveys, focus groups
Evaluation	Experiences, perceptions
Research type	Qualitative

Search strategy

Inclusion and exclusion criteria

As outlined in Table 6.2, the inclusion criteria differed for the qualitative review. Consistent with the effectiveness review, the criteria were structured hierarchically and assessed in order of importance. Included studies needed to use qualitative methods (e.g. interviews, focus groups etc), include perceptions of SIB practitioners on their reasons for adopting SIBs, be published in English and be retrievable. As was described in Chapter 4, in considering if a study reported data on the factors affecting the uptake of SIBs, it was necessary to take a broad approach, as few studies had an explicit objective of understanding why SIBs have been pursued (despite often featuring valuable content related to this). Additionally, it must be noted that motivations are not always explicit and can be very difficult to observe and study. Thus, for this review, if studies looked at factors such as the benefits, criqiques and/or active ingredients of SIBs (and in doing so, broadly reflected on motivations for adoption), they were deemed to be relevant and eligible for inclusion.

The term 'SIB practitioner' was used loosely to refer to stakeholders with any form of involvement in SIBs (e.g. as commissioners, service providers, investors, policy makers, evaluators etc). It was crucial to use a broad definition, given that there are multiple types of SIB actors with diverse interests and this diversity likely accounts, to some extent, for the spread or non-spread of the SIB idea. By allowing for a diverse array of SIB actors to be included in the study samples, it was possible to observe the ways in which the SIB concept, as a 'chameleonic' idea, has been framed in different ways, to adapt to different stakeholder perceptions about what is legitimate.

Table 6.2: Inclusion and Exclusion Criteria – Qualitative Review

Inclusion Criteria	*Exclusion Criteria*
1. Qualitative studies (i.e. studies which use techniques such as in-depth interviews, focus groups, observation, reflective diaries and/or case study methodologies to explore perceptions of practitioners and recipients). 2. Providing evidence regarding the perceptions of practitioners on their reasons for adopting SIBs. 3. May include qualitative evidence collected within studies included in the effectiveness part of the review or relevant stand-alone qualitative studies. 4. Published in English 5. Retrievable	1. Wrong study design 2. Not focused on SIBs 3. Does not include perceptions of practitioners or service recipients regarding the research question 4. Not published in English 5. Not retrievable

Search methods for the identification of studies

The search was conducted simultaneously for the effectiveness review and qualitative review. Thus, the methods regarding the search have already been discussed (Chapter 5) and will not be repeated in this section.

Extraction of information

Study screening and selection

Study screening was also carried out simultaneously for both reviews; thus, the key steps have already been discussed. However, the approach for selecting studies differed and is thus outlined in this section. In a qualitative evidence synthesis, having excess data due to a large number of eligible studies can undermine a researcher's ability to perform a thorough analysis (Ames et al., 2019). Thus, I considered the various approaches for case selection in qualitative research. For instance, Gerring (2008: 1) discusses nine such methods for purposive (non-random) selection of cases, including: *"typical, diverse, extreme, deviant, influential, crucial, pathway, most-similar, and most-different".* Likewise, Suri (2011) proposes a range of strategies; these are adapted from the list produced by Patton (1990) for sampling in primary qualitative research. For this study, four types of sampling were applied

to select appropriate examples from the overall sampling frame, as outlined in Table 6.3. Three of these were used in a related qualitative systematic review, which analyzed the factors affecting evidence-based policymaking in public health (Verboom et al., 2016). Building on this, I also incorporated the 'politically important' sampling approach, which is particularly relevant to the SIBs context, noting the strong political attention that the concept has attracted since 2010.

In brief, I applied four distinct phases of sampling to select studies from the overall sampling frame. I firstly used intensity sampling, which involved selecting studies that appeared to be of a high methodological quality and to contain rich data on factors affecting the uptake of SIBs. Secondly, I used maximum variation sampling, to diversify the sample by adding studies which had characteristics not featured in the first round of sampling. Thirdly, I used confirming and disconfirming cases sampling, to include studies which confirmed or disproved the patterns that had already emerged; in some cases, this involved adding studies that had surprising insights, thus requiring modifications to the synthesis findings. Lastly, and particularly importantly, in the context of SIBs, I sampled politically important cases, to ensure that 'hot' cases had been included. I included a mix of SIBs that were politically 'hot' due to their perceived success (e.g. NSW Newpin SIB), along with projects that have been particularly high-profile and/or attracted polarizing views about whether or not they were successful (e..g Peterborough), and projects that have attracted attention for undesirable reasons (e.g. London Homelessness SIB, in which it was announced that investors were profiting from individuals being deported from the UK). The sampling frame was thus decided upon by drawing on approaches used in existing reviews focused on the use of evidence in policy-making (a closely related topic to this review, which seeks to understand

why SIBs have spread absent evidence of success); and also, making iterations, to account for the politicially sensitive nature of the SIB concept and individual SIB-funded projects.

Table 6.3: Sampling Types Used for Qualitative Analysis

Sampling Type	*Description*	*When Useful/Not Useful*
Intensity sampling	Selecting studies which are *"excellent or rich examples of the phenomenon of interest, but not highly unusual cases"* (Patton 2002: 234).	To develop a comprehensive understanding of many policy changes, it is essential to examine cases where the changes have been embedded in the relevant system over a sufficient time period. Due to factors such as cost, an innovation might be implemented with sufficient intensity in only a few studies, making these studies particularly important.
Maximum Variation (Heterogeneity) Sampling	Identifying unique or diverse variations that have emerged in adapting to different conditions.	Can uncover important insights into a phenomenon by investigating it from all angles. Assists with identifying essential and variable features of a phenomenon as experienced by different stakeholders in different contexts. This can help to facilitate informed global decision-making; and can also help with identifying common patterns that cut across variations.
Confirming and Disconfirming Cases	Involves selecting additional cases which confirm or disprove already emergent patterns in the data.	Important for elaborating and deepening initial analysis, seeking exceptions and testing variation. Confirmatory cases can confirm and elaborate the findings and add richness, depth and credibility. Conversely, disconfirming cases can provide critical rival interpretations and place boundaries around the confirmed findings.
Sampling Politically Important Cases	Involves *"selecting (or sometimes avoiding) a politically sensitive site or unit of analysis"* (Patton, 2002: 241).	Can attract attention to the study (or avoid undesired attention by eliminating potentially politically sensitive cases). Using this approach (and thus synthesizing 'hot topics') can assist with attracting funding and also increase the likelihood of the findings being used and having an impact (Elmore, 1991).

Data extraction and quality assessment of included qualitative studies

While there is no universally accepted approach for extracting qualitative evidence, I created a data extraction form to assist with the process (Appendix F). I used guidance from the EPPI Centre, Cochrane Collaboration and National Institute for Health and Care Excellence (NICE). Data were extracted on factors such as study aims, study design, sample characteristics, theoretical background and findings (Higgins et al., 2019). Qualitative data were then coded using NVivo software. Although the aim of a thematic synthesis is not to determine an effect size, it is still important to consider study quality. Studies were not excluded based on quality, although I did conduct a detailed quality appraisal (Appendix I) and sensitivity testing (Appendix L). There are well-documented limitations to prescriptively

using checklists or scales (Higgins et al., 2019. Regardless, appraisal tools can help to ensure that reviewers treat each study consistently (Gough et al., 2012). Thus, I used a critical appraisal tool developed specifically for qualitative research studies (CASP, 2018; Hannes, 2011; Tong et al., 2007). Specifically, I used the CASP checklist to assess the following for each study: 1) whether the results of the study are valid; 2) what the results are; and 3) whether the results will help locally (i.e. in my specific setting) (Melnyk & Fineout-Overholt, 2015). In answering the 10 questions contained in the CASP checklist, I also referred to more detailed qualitative guides, such as the 'COnsolidated criteria for REporting Qualitative research' (COREQ) checklist, as required (Tong et al., 2007). The quality assessment results for the qualitative studies were also presented in tabular format (Appendix I).

After completing the above critical appraisal steps, I also checked the robustness of the qualitative synthesis. This involved creating a grid (Appendix J) which depicts the relative contribution of each study to the synthesis, as suggested by Rees et al (2009). Importantly, this process helped to identify: 1) whether important parts of the synthesis depend upon few studies; 2) whether this matters; and 3) whether there are associations between specific populations and/or quality judgements of studies and groups of themes. If an area of the synthesis is based on more studies than others, it does not necessarily mean it is more trustworthy, however it is important to be aware of this, when interpreting results.

Additionally, I used the Confidence in the Evidence from Reviews of Qualitative research (GRADE-CERQual) approach for qualitative syntheses (Guyatt et al., 2011) to determine how much confidence to place in each finding (Lewin et al., 2015). For each finding, this involved assessing: 1) methodological limitations; 2) coherence; 3) relevance;

and 4) adequacy of data (Table 6.4) (Lewin et al., 2015). Based on this, I assigned one of the following categories to each of the findings: ‘high’; ‘moderate’; ‘low’; or ‘very low’.

Appendix K shows how the CERQual score (for each component and overall) was arrived at, for each finding; the scores were pooled from each study contributing to a finding, to arrive at the overall scores. Several appendix items also contain information relating to a component of the CERQual assessment; for instance, appendix I (CASP critical appraisal scores) contributed to the “methodological limitations” score; and appendix J provides some insights regarding the adequacy of the data, as it notes the extent to which each study contributed to identified themes.

Table 6.4: GRADE-CERQual Assessment Components

Component	Definition
Methodological limitations	The extent to which there are problems in the design or implementation of the primary studies contributing evidence to the review finding.
Relevance	The extent to which the evidence from the primary studies supporting a review finding is relevant to the review question context (perspective or population, phenomenon of interest, setting).
Coherence	The extent to which a review finding is well grounded in data from the contributing primary studies and provides a convincing explanation for the patterns found in these data.
Adequacy of data	An overall judgement of the quantity and degree of richness of data supporting a review finding.

Source: Guyatt et al., (2011)

Qualitative synthesis

For the review of qualitative evidence, the synthesis was thematic. The analysis involved the following three stages outlined by Thomas and Harden (2008); 1) coding text; 2) developing descriptive themes; and 3) generating broader analytical themes. Each of these steps are briefly discussed below.

Coding text

Firstly, I read each included study in detail and conducted line-by-line (inductive) coding on the findings. A key challenge was deciding what constitutes the 'findings' of included studies. Various strategies exist, which range from being very selective to very inclusive. In inclusive approaches, the entirety of a publication is treated as data; conversely, in selective approaches, data are only considered relevant when explicitly supported by direct quotations from study participants (Noyes & Lewin, 2011). Selective approaches risk missing findings which may be of secondary importance in individual studies (and thus not always supported by direct quotations) but may collectively emerge as essential themes when findings are combined from multiple studies. Additionally, even if a selective approach is taken (e.g. only including quotations from participants), the 'findings' could still reflect the authors' original perspectives and priorities, as quotations may be used selectively to justify their viewpoints. As I was interested in understanding the full breadth of factors that have influenced the uptake of SIBs, I took an inclusive approach and treated the entire paper as 'data'. I thus carefully read each sentence and paragraph to identify underpinning themes. The relevant text was highlighted and associated with a particular code using NVivo. In coding each additional study, I looked at the existing 'bank' of codes I had developed and created new ones where necessary.

Developing descriptive themes

Secondly, I compared and contrasted codes across studies to develop descriptive themes. This step aimed to identify and articulate relationships between themes and to associate conceptually similar themes with each other (Thomas & Harden, 2008). It involved an iterative process in which I looked for similarities and differences between the identified

codes and grouped them hierarchically. It resulted in a tree structure containing several layers of descriptive themes.

Generating broader analytical themes

In the preceding steps, the synthesis did not deviate significantly from the original findings of included studies. The final step involved generating analytical themes that progress beyond the included studies' content, thus offering new conceptualizations and broader explanatory concepts. The notion of 'going beyond' the original content of included studies has been highlighted as the defining characteristic of this type of synthesis (Britten et al., 2002; Thorne et al., 2004; Thomas & Harden, 2008). However, it is also *"the most difficult to describe and is, potentially, the most controversial, since it is dependent on the judgement and insights of the reviewers"* (Thomas & Harden 2008, p. 7). In a meta-ethnography, the equivalent stage is known as the development of 'third order interpretations' that build on the content of original studies (Britten et al., 2002; Campbell et al., 2003). In this step, I critically examined the aggregated descriptive themes developed in stage two and used these to construct analytical groupings representing higher-level findings. This process involved reconsidering the entire dataset, interrogating it for emerging cross-study themes, and building on the descriptive themes to devise analytical groupings. As this iterative process developed, I interpreted each of the emerging themes as factors affecting the uptake of SIBs (Table 6.2) and updated the inventory of factors iteratively as new findings emerged.

Results

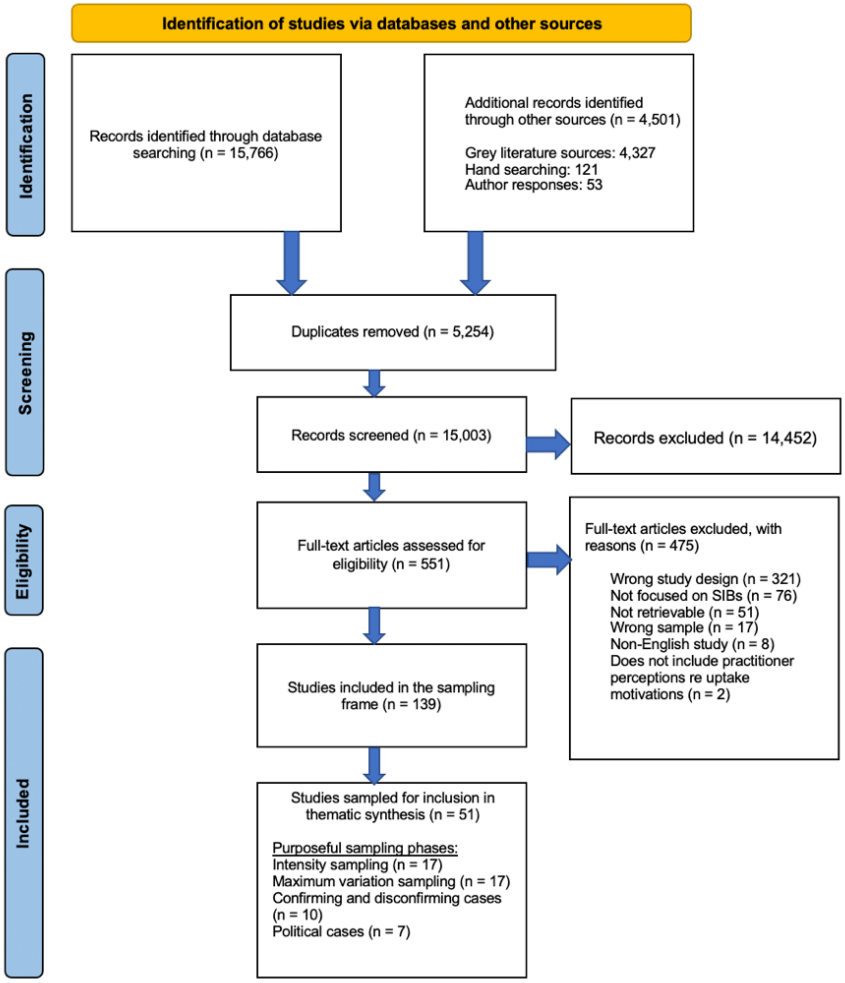
The synthesis of qualitative research sheds light on how broader institutional motives have contributed to the "wide" but "thin" spread of SIBs. The identified themes and sub-themes were linked to the different pillars of legitimacy, thus enabling an exploration of how

SIBs have been adopted as a way of behaving as expected by external constituencies (Suchman, 1995). I present the results using a table (Table 6.2) which outlines the main factors influencing the uptake of SIBs, as identified in the included studies. This involves a complex interplay of efficiency and legitimacy pressures, which have remained largely constant over time (Table 6.3). Where appropriate, I provide explanatory quotations to support the factors identified.

6.3 Findings

The searches of electronic databases, grey literature sources and hand searches were initially conducted in July 2018; and updated between 24-25 January 2021. The PRISMA flowchart presented in Figure 6.1 depicts the flow of studies through the various stages of screening, study selection and sampling (including the various stages of sampling for the thematic synthesis). In total, 15,766 citations were located through electronic database searches. A further 4,501 citations were collected from other sources. After removing duplicates (5,254), 15,003 articles were retrieved for title and abstract review. Of these, 14,452 of these articles were excluded on title and abstract; and 551 were retrieved for full-text review. Finally, 139 were included in the sampling frame and 475 were excluded. Of the 139 studies included in the sampling frame, 51 were sampled for this thematic synthesis (17 from intensity sampling; 17 from maximum variation sampling; 10 through confirming and disconfirming cases sampling; and 7 through politically important cases).

Figure 6.1: PRISMA Diagram – Thematic Synthesis



6.3.1 Characteristics of included studies

In total, 181 studies met the inclusion criteria, and of these, 51 were included in this thematic synthesis. The majority of the included studies primarily used qualitative methods for data collection and analysis (interviews, round tables or case studies). A small number of reports used mixed methods, containing interviews and surveys (Ronicle et al., 2014; KPMG, 2014; Ronicle et al., 2016; ICF Consulting Services, 2019; Loveless, 2020). However, quantitative survey data were irrelevant for this thematic synthesis and thus were not analyzed in this review. Of the 51 studies, 9 included a literature review and qualitative interviews; however, reviews did not meet the inclusion criteria. Thus this portion of the

studies was not included. All but three of the studies in the sample included individual interviews – most commonly with a mix of service providers, investors and commissioners. Samples often also included intermediaries and policymakers. Several of the studies (n = 10) incorporated other qualitative methods (including focus groups, round tables, workshops etc). Table 6.5 presents the characteristics of included studies by different domains.

Table 6.5: Characteristics of included studies – by domain

Domain	Category	N	%
Year of publication	2000 - 2013	5	9.80%
	2014 - 2017	35	68.63%
	2018 - 2020	11	21.57%
Study type	SIB evaluation report	21	41.18%
	Practitioner report	17	33.33%
	Published journal article	9	17.65%
	Research report (non-journal)	1	1.96%
	Roundtable summary	2	3.92%
	Conference paper	1	1.96%
Countries most frequently investigated	UK	29	56.86%
	USA	6	11.76%
	Australia	3	5.88%
	Multiple (USA and UK)	2	3.92%
	Europe	2	3.92%
	Netherlands	2	3.92%
	Canada	2	3.92%
	Cameroon	1	1.96%
	Multiple (Australia, Europe and USA)	1	1.96%
	Multiple (Canada, US, UK)	1	1.96%
	Multiple (Cameroon, India)	1	1.96%
	Unspecified	1	1.96%
Policy areas investigated	Multiple policy areas	12	23.53%
	Child and family welfare	8	15.69%
	Homelessness	7	13.73%
	Unspecified	7	13.73%

	Health	7	13.73%
	Criminal justice	5	9.80%
	Employment & training	4	7.84%
	Education & early years	1	1.96%
Reports written by	Academics or individuals with academic affiliations	19	37.25%
	Consulting firms	19	37.25%
	Independent think tanks	6	11.76%
	Research/evaluation centre	4	7.84%
	Non-profit or charitable organizations	2	3.92%
	Government	1	1.96%
Sample size	Unspecified	24	47.06%
	1-5	1	1.96%
	6-10	1	1.96%
	11-20	5	9.80%
	21-30	7	13.73%
	31-50	8	15.69%
	50-100	1	1.96%
	100+	4	7.84%
Sample type	Service providers	28	54.90%
	Investors	27	52.94%
	Commissioners	26	50.98%
	Unspecified	14	27.45%
	Intermediaries	14	27.45%
	Policymakers	10	19.61%
	Beneficiaries	8	15.69%
	Evaluators	6	11.76%
	Academics	1	1.96%

Regarding publication type, SIB evaluation reports were the most common type. Practitioner reports and published journal articles also featured highly, with only two policy roundtable summaries (CRESSI, 2016; Rotheroe, 2014), a single conference paper (Cooper et al., 2013) and a single non-academic research report (Williams, 2019). In terms of study context, most of the studies focused on a single country setting, with just five studies

focusing on more than one country, often in a comparative manner (Warner, 2013; Rudd et al., 2013; Ormiston et al., 2015; Wainer, 2018; Williams, 2019). The UK was the most well-represented jurisdiction, which reflects the high number of SIBs launched in the UK (n = 93) compared to other countries (with 28 SIBs being the highest number launched in another country (GO Lab, 2023a). While 38 countries have launched SIBs, only seven countries featured explicitly in the studies, thus severely limiting the generalizability of the findings.

A broad array of policy areas were discussed in the included studies. Policy areas in which a high number of SIBs have been launched featured in the studies (e.g. employment and training, child and family welfare, health, education, and homelessness). However, some policy areas (e.g. poverty reduction and agriculture and environment) were completely overlooked. Study samples included a mix of stakeholders involved with SIBs, such as service providers, investors, intermediaries and outcome funders. Particular stakeholder groups (e.g. evaluators and academics) were under-represented in the included studies, while other groups (service providers, investors and commissioners) were well-represented.

Regarding sample size, the average number of study participants (calculated from those studies that specified the sample size) was 56.41 (1,523/27). This thematic synthesis includes studies representing the perspectives of at least 1,523 participants (unclear how many more due to 24 studies not specifying the sample size). When considering the characteristics of included studies, it is also integral to consider whom the relevant studies were written by. In the emerging field of SIBs, there are various stakeholders with vested interests for the innovation to succeed, leading to possible biases in reporting results. A high proportion of the studies were written by consultants, providing further evidence of individuals in the "fashion setting network" playing an active role in disseminating SIBs.

Few of the included studies were published in the early years of the diffusion cycle; the highest proportion were published between 2014-2017. The over-representation of particular years and periods reflects the growth and decline of SIBs interest and usage more broadly.

Table 6.6 provides a summary of the characteristics of each of the 51 studies included in this thematic synthesis (including the sampling frame used for each study). The studies are listed in order of publication year. To assist with ease of reporting results, each of the included studies has been assigned a unique ID number.

Table 6.6: Characteristics of Included Studies – Thematic Synthesis

ID	Citation	Publication year	Study type	Sampling frame	Who written by	Countries	Policy Area(s)	Design	Data Sources	Participant type(s)	No. of Participants	Quality rating
1	Social Finance (2011)	2011	Practitioner report	Politically important	Consultants	UK	Criminal justice	Qualitative	Interviews	Service providers, beneficiaries, commissioner	Did not specify	Low
2	Tomkinson (2012)	2012	Practitioner report	Maximum variation	Academic	Australia	OOHC	Qualitative	Stakeholder perspectives collected during Social Finance Forum	All organizations involved in NSW SIB trial; broader government & investment representatives; Social Finance UK representatives.	Did not specify	Low
3	Warner (2013)	2013	Journal article	Confirming and disconfirming	Academic	USA and UK	Recidivism; early childhood education	Qualitative	Interviews	Investors, policymakers, service providers	Did not specify	Medium
4	Rudd et al., (2013)	2013	Practitioner Report	Confirming and disconfirming	Non-profit	USA and UK	Recidivism	Qualitative	Interviews	Stakeholders involved in the New York City SIB (investors, service providers, commissioners, evaluator)	Did not specify	Low
5	Cooper et al., (2013)	2013	Conference paper	Confirming and disconfirming	Academic	UK	Homelessness	Qualitative	Archival (public documents from government, NGO, and corporate sources) and interview data	2 interviews total, with 3 practitioners (an investor and staff from Greater London Authority) (one interview had 2 people in it)	3	Low
6	Ronicle et al., (2014)	2014	Evaluation report	Maximum variation	Consultants	Unspecified	Unspecified	Mixed methods (survey and interviews)	Review of available literature about SIBs (all literature published up to 20 June 2014 and two important evaluations of SIBs published later in Summer 2014); a series of in-depth consultations with organizations centrally involved in the policy, strategic and/or operational development	Commissioners, investors, policymakers and service providers.	Telephone and face-to-face consultations undertaken with eight organizations centrally involved in the policy development, strategic development	Medium

									of SIBs; and surveys with the core groups involved in SIBs (commissioners, investors and service providers).		and operational delivery of SIBs in the UK. Tele surveys undertaken with 19 investors, and e-surveys undertaken with 24 commissioners and 49 service providers either involved in or very informed about the SIB agenda.	
7	Hughes and Scherer (2014)	2014	Practitioner Report	Confirming and disconfirming	Consultants	USA	Unspecified	Qualitative	Interviews	Thought leaders at 30 different foundations and organizations involved in impact investment	Did not specify	Low
8	Wallace and Roxas (2014)	2014	Evaluation report	Politically important	Research/evaluation	Australia	OOHC	Qualitative	Group discussions and a small number of one-on-one interviews	Newpin program management representatives including the Newpin Operations Manager, the Newpin Practice Coordinator, the UnitingCare Children, Young People and Families Manager of the Research and Development, and the UnitingCare Children, Young People and Families Director of Business of Development and Finance. Newpin staff, including the Newpin Manager, Coordinators, family workers, and play staff from all four Newpin Centres in western	Did not specify	Medium

										Sydney (21 staff in total), the FACS Newpin Contract Manager and the Principal Policy Officer, Funding and Contracting.		
9	Disley and Rubin (2014)	2014	Evaluation report	Politically important	Research/evaluation	UK	Recidivism	Qualitative	Semi-structured interviews	Practitioners involved with Peterborough SIB (organizations/roles not specified)	39	Medium
10	KPMG (2014)	2014	Evaluation report	Maximum variation	Consultants	Australia	OOHC	Mixed methods (survey and interviews)	Interviews; survey; review of literature; transaction costs data; survey	Practitioners involved in SIB trial (policymakers, commissioners, service providers, investors, intermediaries)	21	High
11	Galley et al., (2014)	2014	Practitioner Report	Maxmim variation	Independant think tank	Canada	Unspecified	Qualitative	Review of existing literature and in-depth interviews	NFP service providers, funders, and policymakers who have direct experience with SIBs	22	Low
12	Mason et al., (2014)	2014	Evaluation report	Politically important	Consultants	UK	Homelessness	Qualitative	Detailed review of a range of documents relating to the development of the SIB and qualitative interviews	Commissioners and strategic stakeholders; London boroughs; the two providers delivering the SIB; their partners and other providers in London; and social investors and intermediaries.	38	Low
13	Rotheroe (2014)	2014	Policy Roundtable Summary	Maximum variation	Consultants	UK	Unspecified	Qualitative	Data collected during a roundtable discussion	Sector experts and charities involved in SIBs	Did not specify	Low
14	Reeder et al., (2014)	2014	Practitioner report	Confirming and disconfirming	Academic	Europe	Unspecified	Qualitative	Interviews	Practitioners from 15 investment firms	Did not specify	Medium
15	Ronicle and Stanworth (2015)	2015	Evaluation report	Maximum variation	Consultants	UK	Health	Qualitative	Interviews	Did not specify	Did not specify	Low
16	Ecorys and ATQ	2015	Evaluation report	Maximum variation	Consultants	UK	OOHC	Qualitative	Interviews	Commissioner - representative/s from North Somerset Council	Did not specify	Low

	Consultants (2015)											
17	Disley et al., (2015)	2015	Evaluation report	Politically important	Research Institute	UK	Recidivism	Qualitative	1) Qualitative, semi-structured interviews with 29 stakeholders; 2) Qualitative, semi-structured interviews with 15 service users of the One Service; 3) a review of the case files of each of the 15 One Service cohort members interviewed; 4) information from the One Service case management database	Practitioners and service users	44	Medium
18	Ormiston et al., (2015)	2015	Journal article	Confirming and disconfirming	Academic	Australia, Europe and the United States	Various sectors	Qualitative	Qualitative case study approach	Representatives from 10 organizations across Australia, Europe and the United States that had implemented impact investment practices.	Did not specify	Medium
19	Reeder et al., (2015)	2015	Journal article	Confirming and disconfirming	Academic	Europe	Unspecified	Qualitative	Interviews	Practitioners from 16 investment firms	Did not specify	Medium
20	Roberts and Cameron (2015)	2015	Evaluation report	Intensity	Consultants	UK	OOHC	Qualitative (interviews)	Interviews; consultations/workshops	MST stakeholders interviewed who are directly involved with Essex SIB. Also, 12 completed the survey.	Did not specify	Medium
21	DCLG (2015)	2015	Evaluation report	Politically important	Government	UK	Homelessness	Qualitative	Interviews	Wide range of stakeholders, provider staff and members of the cohort of entrenched rough sleepers in receipt of support. The sample included 49 stakeholders, a mix of commissioners, investors, providers, and beneficiaries.	49	Medium

22	Tan et al., (2015)	2015	Evaluation report	Intensity	Academic	UK	Health and social care	Qualitative (semi-structured interviews)	Literature review + semi-structured interviews	Mix of service providers, commissioners, investors, intermediaries, and others involved in the SIB (e.g. Cabinet Office staff)	36 total, including service providers (n=11), commissioners (n=6), investors (n=2), intermediaries (n=11) and others involved in SIBs (n=6), e.g. staff at the Cabinet Office	High
23	Sandstrom and Greenberg (2016)	2016	Practitioner report	Intensity	Think tank	USA	Early childhood education	Qualitative	Hour-long interviews	Government leaders, evaluators, and intermediaries from projects at four different sites.	6 SIB stakeholders from projects at four different sites (24 total)	Medium
24	Ecorys UK and ATQ Consultants (2016a)	2016	Evaluation report	Maximum variation	Consultants	UK	Workforce development	Qualitative	In depth, semi-structured interviews + focus group	Interviews with two commissioners: Staffordshire County Council (Staffordshire), London Borough of Haringey (Haringey); three providers of services – Making Space, Twining Enterprise and Working Well; investors/outcomes payers: Big Issue Invest; DCMS - Social Outcomes Fund; and The National Lottery Community Fund - CBO; and MHEP team. Focus group with Working Well Trust's IPS service team.	Did not specify	Low
25	Van Es et al., (2016)	2016	Practitioner report	Maximum variation	Academic	Netherlands	Youth unemployment	Qualitative	Wide-ranging literature review, supplemented by three semi-structured interviews and a written	Semi-structured interviews undertaken with respondents working for the municipality of	Did not specify	Medium

									exchange of information with experts	Rotterdam (as a government organization) and investor ABN AMRO Bank. Written exchange of information undertaken with the social service provider (The Buzinezzclub) who could not participate in an interview due to time constraints. The third interview was conducted with the project leader (Deloitte) of a SIB that was being implemented at the time of the interview.		
26	Ronicle et al., (2016)	2016	Evaluation report	Maximum variation	Consultants	UK	Various - health, kids in OOHC	Mixed methods (survey and interviews)	Literature review; stakeholder consultations/interviews surveys; in-depth review reports; case study report	Investors, commissioners and service providers completed the survey; consultations were undertaken with 'organizations centrally involved in the policy development, strategic development and operational delivery of SIBs in England'	Did not specify	Low
27	Griffiths et al., (2016)	2016	Evaluation report	Intensity	Consultants	UK	Youth NEET	Qualitative	Semi-structured face-to-face interviews	Mix of investors, intermediaries, delivery managers, schools, frontline staff, and beneficiaries (young people) involved with IF projects.	204 face-to-face interviews were conducted across the ten IF projects. In total, 104 interviews were conducted with key stakeholders: 18 with investors, 11 with intermediaries, 24 with delivery managers, 25	Medium

											with schools, 24 with frontline, client-facing staff and two with other stakeholders. Approximately half were repeat interviews with stakeholders first interviewed in 2012. A total of 100 young people who had participated in Round One and Round Two IF projects were also interviewed.	
28	James, Kane and Ravenscroft (2016)	2016	Practitioner report	Intensity	Consultants	UK	Homelessness	Qualitative	Three strands of work: 1) an analysis of the structure and finances of social sector organizations examining the sector's existing financing arrangements and their assets and liabilities; 2) a literature review on organizations' financing behaviours from research that has already taken place on the UK social investment market; 3) case studies on organizations' social investment journeys.	4 organizations, including Depaul - an organization with a SIB contract worth more than £1m	Did not specify	Low
29	Ecorys UK and ATQ Consultants (2016b)	2016	Evaluation report	Maximum variation	Consultants	UK	Health - social isolation	Qualitative	Review of documents provided by stakeholders and consultations with key	"Key stakeholders involved in the SIB, including representatives from the commissioners,	Did not specify	Low

									stakeholders involved in the SIB, including representatives from the commissioners, intermediaries, providers and investor	intermediaries, providers and investors"		
30	Cooper et al., (2016)	2016	Journal article	Confirming and disconfirming	Academic	UK	Homelessness	Qualitative	Multiple interviews + analysis of documentary evidence on St Mungo's, SIBs and the London Homeless SIB	The major parties involved in the construction of and implementation of the SIB including five senior managers from the three key organizations involved in the funding arrangement, along with other figures from the voluntary/charitable sector. We also	Did not specify	Medium
31	CRESSI (2016)	2016	Policy Roundtable Report	Maximum variation	Academic	UK	Various - Youth NEET; OOHC	Qualitative (other)	Roundtable discussion	Commissioners, academics, policymakers, investors, service providers	32	Medium
32	Smeets (2017)	2017	Journal article	Intensity	Academic	Netherlands	Workforce development	Qualitative (Interviews and focus group)	Interviews and a focus group	3 preliminary interviews (unspecified sample); followed by interviews with the 12 key persons of this SIB; focus group consisting of six participants (from five organizations)	21	Medium
33	Moffatt et al., (2017)	2017	Journal article	Intensity	Academic	UK	Health	Qualitative study	Semi-structured interviews	Service beneficiaries	Thirty adults with long-term conditions, 14 females, 16 males aged 40–74 years, mean age of 62 years, 24 white British, 1 white Irish, 5 from black and	High

											minority ethnic communities.	
34	Laing et al., (2017)	2017	Evaluation report	Intensity	Academic	UK	Health	Qualitative	Semi-structured interviews and focus groups with link workers; semi-structured interviews with patients referred to WtW, examining their experiences of engaging with the service	Mix of service providers and beneficiaries	45 total. First phase: five focus groups and fifteen individual interviews were undertaken; second phase, four focus groups with link workers and senior link workers were undertaken. Fifteen link workers participated in the first wave of data collection (five focus groups and 15 interviews); and 15 in the second wave (four focus groups). Thirty WtW clients were interviewed as part of this research.	Medium
35	Hawkins and Bieretz (2017)	2017	Practitioner report	Maximum variation	Think tank	USA	Various	Qualitative	Interviews plus a review of existing PFS projects contracts and procurement documents	Practitioners from six projects across five locations: Cuyahoga County, Ohio; Denver, Colorado; Santa Clara County, California; South Carolina; and Connecticut. In each location, 1-3 people involved with procuring	Did not specify	Low

										or structuring the project were interviewed. Whenever possible, the sample included the government official who helped design the project and select the partners. The researchers also spoke to representatives from two professional procurement associations about their assumptions regarding PFS and procurement.		
36	Hawkins et al., (2017)	2017	Practitioner report	Politically important	Think tank	USA	Recidivism	Qualitative	Comprehensive landscape scan of all PFS efforts with a criminal justice component. Information was found on the websites of various PFS stakeholder organizations. The authors then reached out to project managers to review project information. Some project managers were also asked for a phone call to discuss criminal justice–related PFS lessons learned. The authors then conducted interviews with 14 individuals, which encompassed 19 PFS efforts in the landscape scan. Interview protocols focused on the motivations of jurisdictions to use PFS for criminal justice, the	Individuals from 19 PFS projects	14	Low

									reasons behind the prevalence of justice-related outcome measures, and lessons learned from the projects.			
37	Neyland (2017)	2017	Journal article	Confirming and disconfirming	Academic	UK	Children at risk of going into care	Qualitative	Semi-structured interviews	Investors, delivery agencies, intermediaries, frontline workers and policymakers for the SIB.	10	Low
38	Walsh et al., (2017)	2017	Practitioner report	Intensity	Think tank	USA	Unspecified	Qualitative	Interviews	Individuals involved in PFS projects, representing seven state or local government agencies, three for-profit investors and three philanthropic investors. Collectively, the group has connections to 15 PFS projects in either the planning or implementation phases.	13	Medium
39	Loxley (2017)	2017	Practitioner report	Maximum variation	Academic	Canada	Child welfare	Qualitative	Interviews	Unclear - mentions interviews with state officials and 'personal interviews'	Did not specify	Low
40	Ronicle et al., (2017)	2017	Practitioner report	Intensity	Consultants	UK	Health and social care	Qualitative	Consultations/interviews with stakeholders, minimal detail provided	Commissioners, service providers and investors across 25 sites	Did not specify	Medium
41	Edmiston and Nicholls (2018)	2018	Journal article	Intensity	Academic	UK	Homelessness; OOHC; youth unemployment	Qualitative	Analysis of stakeholder documentation, process and impact evaluations, and qualitative fieldwork with cross-sectoral stakeholders.	Policymakers, public sector commissioners from local and central government, third sector service providers, social finance intermediaries and entities, social investors and service users. The majority of these participants were directly involved in (or affected by) the	41	Medium

										financing, governance and service interventions delivered through the four SIBs considered. A smaller number of national-level stakeholders were also interviewed who were more tangentially involved but have nonetheless been instrumental in the development and uptake of SIBs at the national level. The researchers also organized a practitioner seminar and a policy roundtable to facilitate an applied discussion between policy-relevant stakeholders and also undertook a range of informal interviews and correspondence.		
42	Tse and Warner (2018)	2018	Journal article	Confirming and disconfirming	Academic	USA	Early childhood services	Qualitative	Review of publicly available contract and loan agreements, press releases, and journalistic articles plus semi-structured interviews	Five interviews for the South Carolina SIB, three for the Chicago SIB, and three for the Utah SIB. Among these interviews, one was a public agency, one was an evaluator, two were funders, two were service providers, three were technical assistance providers, and two were intermediaries. Interviewees were selected based on their involvement in the design, launch, and	11	Medium

										implementation of the SIB.		
43	Oroxom et al., (2018)	2018	Practitioner report	Maximum variation	Think tank	Cameroon	Health	Qualitative	Interviews plus a review of the mostly grey and published literature related to DIBs and SIBs	Staff from the organizations involved with the cataract bond and individuals with knowledge of innovative financing more broadly	Did not specify	Low
44	Fraser et al., (2018)	2018	Evaluation report	Intensity	Academic	Evaluation report	UK	Health and social care	Qualitative case study comparisons, including semi-structured interviews, SIB contract analyses; quantitative analysis was planned but could not be undertaken due to data limitations.	A mix of commissioners (n=38 with 32 informants), providers (n=123 with 109 informants), SIB specialist organizations (n=23 with 13 informants), investors (n=9 with 10 informants) and others (n=5), e.g. central government, or data analysts or consultants, or independent board members.	177 interviews undertaken with 199 informants overall across all sites between June 2014 and May 2017 until 'data saturation.'	High
45	Wainer (2018)	2018	Practitioner report	Maximum variation	Charity	Cameroon, India are the focus of the case studies (UK, USA, Ethiopia also mentioned)	Health	Qualitative	Two case studies of DIB projects. Data sources included: 1) a review of relevant secondary documents; and 2) in-depth, semi-structured interviews	Eight stakeholders were interviewed for the Rajasthan maternal care DIB, and six were interviewed for the Cameroon Newborn DIB.	14	Low
46	Williams (2019)	2019	Research report	Intensity	Academic	Canada, USA, UK	The scope is limited to social services with a particular emphasis on five core areas: (1) employment; (2) early education; (3) homelessness; (4) reoffending; and (5) child	Qualitative	1) Review of publications, reports, and blogs on the subject of SIBs, PFS, and outcomes-based funding; 2) in-depth interviews.	Most of the interviews were undertaken with individuals who had either directly participated in a SIB or had engaged in exploratory or feasibility work. This included interviewees selected from the following groups: (1) government;	196 (Canada, n = 54; US, n = 68, UK, n = 74)	Medium

							welfare. Passing references are made to health SIBs as well as environmental and DIBs.			(2) providers; (3) investors (including trusts and foundations); (4) SIB specialists (intermediaries, advisors, and fund managers); and (5) evaluators and data specialists. Additionally, given the interest in the world beyond SIBs and how this funding model fits into the larger non-profit funding ecosystem, the researcher also interviewed a larger group of social service providers, philanthropists, social and impact investors, philanthropic consultants and non-profit advisors, and those engaged in policy and advocacy work.		
47	ICF Consulting Services (2019)	2019	Evaluation report	Intensity	Consultants	UK	Homelessness	Mixed methods	Program performance data; qualitative costed case studies (involving face-to-face interviews with a small sample of participants; interviews with their key workers; and reviewing relevant data); and in-depth interviews and focus groups.	Mix of service providers (n=108); local authority staff (n = 26); staff involved with referrals (n = 18); investors and intermediaries (n = 18); and young people (year 1, n = 70; year 2, n = 35; year 3, n = 20).	277 (although unclear if the same young people were interviewed in years 1, 2, 3)	Low
48	Wooldridge et al., (2019)	2019	Practitioner report	Intensity	Consultants	UK	Employment and training, health, homelessness, child and family welfare,	Qualitative	Rapid evidence assessment; case studies; interviews; workshop	Interviews were conducted with 'key sector experts'; a workshop was conducted with a mix of local commissioners, central	17 (interviews); 25 (workshop)	Low

							education, criminal justice			government commissioners, sector experts, and policymakers.		
49	Loveless (2020)	2020	Evaluation report	Intensity	Consultants	UK	Child and family welfare, employment and training, health, homelessness, education, criminal justice	Mixed methods	Administrative project-level data (held by TNLCF); Qualitative interviews; Case study fieldwork (including additional interviews) with three successful Round 1 applicants whose projects were in the early stages of implementation	30 for qualitative investigation; unspecified for case study fieldwork		Medium
50	Rosenback and Carter (2020)	2020	Evaluation report	Intensity	GO Lab - research centre	UK	Homelessness	Qualitative	Logic models developed with frontline staff and provider organizations; semi-structured interviews conducted	Council contract managers, provider managers, the social investment fund manager and referral partners from local voluntary sector organizations and public sector social care and housing representatives.	23	Medium
51	Stanworth (2020)	2020	Evaluation report	Maximum variation	Consultants	UK	Health	Qualitative	Interviews	Key stakeholders that represent all the main parties to the SIB, including: 1) The EJAF Chief Executive, Head of Projects and day-to-day Performance Manager for this SIB; 2) The lead commissioner (LB Lambeth), which represents all three of the LSL Boroughs on the Zero HIV CIC Board; 3) Three of the six providers, representing each of the main settings in which the services are delivered - i.e. a Hospital	Did not specify	Low

										Trust, a GP Foundation and a community-based VCSE; 4) All the investors except ViiV, whose representative had recently left the company and not yet been replaced; and 5) The National Lottery Community Fund as administrators and payers of the CBO grants to EJAF.		

6.3.2 Quality appraisal

Included studies were appraised for quality (Appendix I), noting that systematic bias can result in inaccurate or untrustworthy findings (Mhasker et al., 2009). Applying the CASP checklist (also informed by COREQ), four studies (7.84%) were rated 'high quality', 24 studies (47.06%) were rated 'medium quality'), and 23 studies (45.1%) were rated ‘low quality’. The CASP rating was one aspect that was taken into account when assessing the CERQual rating (i.e. level of confidence) for each finding (Appendix K).

6.3.3 Excluded studies

At full-text review, 463 studies were excluded. Articles were primarily excluded for having the wrong study design (n = 321). Other exclusion reasons included; not focusing on SIBs (n = 76); not being retrievable (n = 51); having a wrong sample (n = 17); full text not being available in English (n = 8); and practitioner perspectives about SIB motivations not being included (n = 2).

6.3.4 Thematic Synthesis: Motivations for Pursuing SIBs

This synthesis uncovered various factors that have affected the rapid and widespread growth of SIBs, as presented in Table 6.7. The “confidence in finding” column shows the CERQual rating for each finding (arrived at by pooling the scores, from each contributing study, as shown in Appendix K). It shows that the level of confidence, for many findings, was assessed as being ‘low’ or ‘very low’, which is reflective of factors such as the low study quality of many of the contributing studies. In the section below, I present the analytical themes related to the motivations, split into the following: 1) factors relating to instrumental motivations; and 2) factors relating to institutional motivations (according to each of the pillars of legitimacy). For the institutional motives, I use the four pillars of legitimacy, which

are well documented in the institutionalist literature, as the overarching framework, for the iterative coding that I conducted (Greenwood et al., 2017). While there are limitations with 'vote counting', I used this approach to quantify the number of studies in which each factor was mentioned (Cwikel et al., 2000; Hedges & Olkin, 1980). Categorizing the factors as either an institutional or instrumental motivation was subjective, and some factors could be dual classified. For the sake of brevity, exemplar quotations are not provided for each identified factor, although the top cited factors are discussed in more detail, below Table 6.7.

Table 6.7: Factors Associated with the Adoption of SIBs

	Theme/finding	*Number of studies cited in (all)*	*Number of studies cited in (2010 – 2013)*	*Number of studies cited in (2014 – 2017)*	*Number of studies cited in (2018 – 2020)*	*Number of studies cited in (low qual removed)*	*Confidence in finding*
INSTRUMENTAL MOTIVATIONS							
1	To improve outcomes	31	4	19	8	15	Low
2	To achieve greater focus on outcomes	35	2	25	8	18	High
3	To achieve greater efficiency	14	1	8	5	8	Moderate
4	To improve service quality	40	3	27	10	22	Moderate
5	To achieve better value for money	13	2	6	5	7	Moderate
6	To achieve savings	30	5	17	8	15	Low
7	To achieve better accountability	30	2	19	9	20	Moderate
8	To foster better measurement/EBI principles	42	3	29	10	25	Low
9	To test the effectiveness of interventions	24	3	13	8	14	Low
10	To test the effectiveness of innovative funding model	9	2	5	2	4	Very low
INSTITUTIONAL MOTIVATIONS							
Legitimacy pillar one: normative							
11	To fill a gap in service provision	35	4	22	9	15	Low
12	To focus on particularly vulnerable groups	20	2	12	6	11	Low
13	To shift the focus to early intervention/prevention	6	1	4	1	3	Low
14	To enhance collaboration	38	2	29	10	23	Moderate
15	To expand services	16	2	11	3	9	Moderate
16	To foster innovation	17	1	9	7	9	High
17	To align with growing/trendy policy area	12	1	9	2	7	Moderate
18	To align with long-term governmental reform	11	1	7	3	1	Moderate

19	To achieve a long-term focus on outcomes	12	1	8	3	8	Moderate
20	To diversify market of providers	3	0	3	0	1	Very low
21	To stimulate competition	20	3	13	4	10	Low
22	To benefit from private markets/private rigour	30	3	19	9	14	Low
23	To scale interventions	28	3	16	9	13	Low
Legitimacy pillar two: pragmatic							
24	To achieve additionality of funding/resources	32	3	20	10	16	Low
25	To achieve competitive advantage for investors	5	0	3	2	4	Low
26	To allow for the recycling of capital	4	1	2	1	4	Very low
27	To obtain investor profits	7	0	5	2	3	Low
28	To transfer risk	45	5	29	12	24	Low
29	To achieve savings for government	30	5	18	7	15	Low
30	To contribute to CSR/charitable goals/achieve alignment to mission and values	6	0	5	1	3	Very low
31	To achieve greater funding certainty for NGOs	25	2	17	6	15	Moderate
32	To avoid upfront costs	19	2	13	4	12	High
33	To blend economic and social value	31	5	20	6	16	Low
34	To pursue neoliberal agenda/motivations	1	1	0	1	1	Very low
Legitimacy pillar three: regulatory							
35	Passing of enabling legislation	5	1	3	1	3	Very low
36	Central government support and policy advice	3	0	0	3	1	Low
37	Impact of existing legislation	2	0	0	2	1	Low
Legitimacy pillar four: cognitive							
38	To build capacity and/or build marketplace for impact investment	29	3	18	8	14	High
39	Evaluation/evidence challenges	39	4	26	10	24	Moderate
40	Harm/unintended consequences	21	3	11	7	12	High
41	No clear blueprint/precedent, findings not 188generalizable etc	25	2	16	7	12	Moderate
42	'SIB effect' unclear	23	2	15	6	11	Low
43	Lack of appropriate knowledge, skills, resources, funding	27	1	21	5	14	Low
44	Cost savings unclear, difficult to calculate, focused on too much	24	4	12	8	12	Low
45	Cautionary academic literature/voices & lack of required literature	13	0	8	5	7	Moderate
46	Difficulties with testing feasibility	12	0	7	5	8	Moderate
47	Added complexity	35	3	25	7	19	Low
48	Confusion over aspects of SIBs contracts/PFS	9	0	4	5	5	Low

How efficiency and legitimacy motivations have changed over time

Understanding how the mix of efficiency and legitimacy drivers have changed over time, as levels of uncertainty have changed, is of central importance to my research question. Thus, in Table 6.8, I demonstrate how the themes have shifted over time. The table demonstrates that the instrumental factors have been most common over time (despite a severe lack of evidence), while most of the legitimacy pillars (normative, pragmatic and cognitive) have also been well-represented across time. The regulatory pillar has barely featured across all time points; it was only mentioned in studies published between 2013-2014 and 2017-2019. Building on this, Appendix L shows that while the ordering of the top 5 factors for each time point changed slightly, a very similar mix of factors was included in the early, middle and later stages of the diffusion cycle. Notably, instrumental motivations have been widely discussed, even when the quantitative evidence base has remained limited, and institutional motivations have been prominent from the outset rather than only in the later stages of adoption. The finding that there is no clear pattern (e.g. 2-stage process) in drivers according to the timing of diffusion is inconsistent with the early institutional literature, which found that early adopters are motivated by technical gains, while later adopters are motivated by social gains (Tolbert & Zucker, 1983). However, it is aligned with more recent literature, which has noted that it can be challenging to disentangle efficiency and legitimacy motivations, which can co-exist (Kennedy & Fiss, 2009).

Table 6.8: Efficiency and Legitimacy Factors Over Time

Year	N (studies)	Instrumental factors (n)	Instrumental factors (%)	Legitimacy (normative) factors (n)	Legitimacy (normative) factors (%)	Legitimacy (pragmatic) factors (n)	Legitimacy (pragmatic) factors (%)	Legitimacy (regulatory) factors (n)	Legitimacy (regulatory) factors (%)	Legitimacy (cognitive) factors (n)	Legitimacy (cognitive) factors (%)
2010	0	0	0%	0	0%	0	0%	0	0%	0	0%
2011	1	1	100%	1	100%	1	100%	0	0%	0	0%
2012	1	1	100%	1	100%	1	100%	0	0%	1	100%
2013	3	3	100%	3	100%	3	100%	1	33.33%	3	100%
2014	9	9	100%	9	100%	8	88.89%	1	11.11%	9	100%
2015	8	8	100%	7	87.5%	7	87.5%	0	0%	8	100%
2016	9	9	100%	9	100%	8	88.89%	0	0%	8	88.89%
2017	9	9	100%	8	88.89%	8	88.89%	2	22.22%	8	88.89%
2018	5	5	100%	5	100%	4	80%	1	20%	5	100%
2019	3	3	100%	3	100%	3	100%	3	100%	3	100%
2020	3	3	100%	3	100%	3	100%	0	0%	2	66.67%

Instrumental motivations

The instrumental category includes themes relating to increased effectiveness and/or efficiency post-adoption. Instrumental drivers were cited in many of the included papers, despite the lack of evidence in the existing literature surrounding the technical benefits of SIBs (see Chapter 5). A frequently cited motivation for pursuing SIBs was that SIBs are, in theory, aligned to an increasing focus on evaluation and data and can thus foster better measurement practices and evidence-based practices (KPMG, 2014; Tan et al., 2015). In one study, the importance of investing in resources for data collection and analysis was stressed by informants (KPMG, 2014). There was a strong sense that while public sector and charitable organizations have traditionally failed to monitor operational and financial performance, SIBs can alleviate these shortcomings through a greater focus on measurement and performance management (Galley et al., 2014; Rudd et al., 2013).

Another frequently cited motivation was achieving cashable savings. It was noted that SIBs are often proposed for programs intended to achieve government savings in a relatively short time frame (Tan et al., 2015; Wallace & Roxas, 2014). In some SIBs, the intention is that any premium paid to investors should be paid either partially or wholly from cashable

savings generated for the commissioner through the achievement of agreed-upon social outcomes (Tomkinson, 2012). While this can be a motivation for launching SIBs, it was noted that SIBs do not necessarily require governments to save money (Ronicle et al., 2014). Various examples were provided in which cashable savings have been an explicit goal. For instance, the Essex and Manchester SIBs both envisaged savings; and the 'It's All About Me' (IAAM) Adoption Bond *"depends on local authority purchasers accepting that their costs will be lower if they buy the service than if the child remained in its current placement"* (Ronicle et al., 2014, p. 30).

SIB stakeholders have expressed excitement at using SIBs to test unproven, novel interventions to address specific policy problems (Ronicle et al., 2014). The included studies found that SIBs are sometimes pursued to create new, localized experiments (Tan et al., 2015). For commissioners, this can be an important motivation as it allows them to *"pilot new service models or defer taking financial responsibility for innovative or risky projects that require testing before scaling up or disseminating"* (Tan et al., 2015, p. 74). Equally, service providers can be drawn to SIBs as the model provides financial freedom and flexibility in service delivery, thus enabling them to explore different preventative care models (CRESSI, 2016; Ecorys & ATQ Consultants, 2015).

Box 1: example (increased efficiency)

"With a FaCS budget of approximately $5 billion per annum, of which over $2 billion per annum funds non-government organizations, even a small spillover benefit of improved contracting efficiency far outweighed the $5 million or so per annum that would be spent on Social Benefit Bonds" (Tomkinson 2012, p. 6).

Institutional motivations

1) Normative legitimacy pillar

This category included themes about whether SIBs are 'right' for society. It was widely cited, particularly in comparison to the regulatory and cognitive pillars, suggesting that the adoption of SIBs has largely been driven by beliefs about SIBs being aligned to the values and norms of society. Likewise, a failure to use SIBs has often been linked to strong perspectives regarding how the funding model is 'wrong' for society (e.g. ethically wrong for investors to profit from helping the most vulnerable individuals etc.). Opposing arguments often contain highly emotional and moralizing rhetoric, focusing on justice and ethics rather than individual gains (Warner, 2013). The three most commonly cited themes (foster innovation, align to a growing/trendy policy area and enhance collaboration) are briefly discussed below.

A commonly mentioned motivation for pursuing SIBs was to foster innovation (CRESSI, 2016). There was a belief, and in some cases evidence, that SIB services were delivered more flexibly than similar non-SIB services, leading to innovation in service delivery (Tan et al., 2015). Specifically, it was mentioned that the commissioners' interest in SIBs was associated with the drive to foster innovation (Tan et al., 2015; Tomkinson, 2012; Cooper et al., 2016; Loxley, 2017; Edmiston & Nicholls, 2018; Ronicle et al., 2014). Similarly, service providers stressed that SIBs enable the use of outcome rather than process measures, thus enabling greater flexibility for service providers in how they deliver services (Disley & Rubin, 2014). Additionally, SIB stakeholders often viewed the flexibility of the SIB model as being particularly advantageous as it fostered local approaches to specific problems, allowing innovation in ways that might otherwise be perceived as being too risky

(e.g. testing the link between social isolation and loneliness with improved health outcomes in Worcester) (Tan et al., 2015).

It was found that SIBs are perceived to be part of a growing and ‘trendy’ policy area, thus making them desirable for governments to pursue (Ronicle et al., 2014). For example, there have been extensive developments in the impact investing landscape, with the market moving from 'uncoordinated innovation' to 'market building' (KPMG, 2014). These developments have been fostered by a growing interest in “double bottom line” investing, characterized by a blend of for-profit investing and the promotion of public good. SIBs are a financial instrument that can provide a mix of social and financial returns to investors (Disley et al., 2015; Neyland, 2017). Interest in SIBs also aligns with the continuous shift towards paying service providers for the outcomes they deliver rather than paying for activities measured by outputs. Broadly, PbR contracts have been widely used in the UK and are a cornerstone of the Government's 'Open Public Services' agenda; thus, SIBs have evolved out of PBR government contracts (Mason et al., 2014). In the UK, SIBs have also been assisted by the emergence of a government-supported social finance market (Galley et al., 2014). It is well recognized that Not-For-Profit organizations (NFPs) were struggling to access funding due to steep funding cuts; the continuous shift towards PBR contracts forced NFPs to carry the financial burden and risk (Oroxom et al., 2018). In this context, SIBs have presented NFPs with an opportunity to relieve financial pressure; rather than taking on significant risk, they have been able to access consistent, multi-year funding through an outcomes-based funding approach (Rudd et al., 2013; Galley et al., 2014).

Enhancing collaboration is also a frequently mentioned motivation; in a SIB contract, effort is required to align the varying interests of different stakeholders (DCLG, 2015;

Wallace & Roxas, 2014; Tomkinson, 2012; Ronicle et al., 2014). SIBs allow for a more direct role to be built between stakeholders, for instance, between philanthropists and governments (Tomkinson, 2012). The successful implementation of a SIB requires a genuine commitment to partnership working between all stakeholders. One of the core strengths of the model is often perceived to be *"its ability to facilitate multisector, multi-partner collaboration to generate both social and financial benefits"* (Hughes & Scherer, 2014, p. 8). In the Peterborough SIB, *"good partnerships between the One Service and the local authority, police, the probation service and HMP Peterborough were reported, both at strategic and operational levels"* (Disley & Rubin, 2014, p. 5). In the London Rough Sleepers SIB, effective working relationships were highlighted as being crucial for project success; the project featured regular meetings, information sharing and a joint commitment to achieving successful outcomes for clients (DCLG, 2015). Similarly, in the Newpin program, the adoption of a collaborative approach (including partnerships between government and NGOs and between NGOs) was seen to be crucial.

Box 2: example (shifting focus to preventative programs)

> *"Several of our interviewees cited SIBs' emphasis on funding preventative programs as an appealing characteristic; as Frederick Douglass once said, "It is easier to build strong children than to repair broken men"*
> (Hughes & Sherer 2014, p. 17)

2) Pragmatic legitimacy pillar

The pragmatic pillar is based on the self-interested calculations of the most immediate stakeholders (e.g. investors, consultants etc) (Suchman, 1995; Walker et al., 2014). Along with the normative pillar, this featured highly in the included studies, suggesting that there is some agreement on the potential benefits of SIBs for particular groups; and that these possible benefits have driven SIB adoption in some instances. The three most commonly

cited themes included: 1) additionality of funding; 2) risk transfer; and 3) achieving savings for government (already discussed).

Several included studies noted that SIBs may deploy a significant new stream of funds by raising investment capital (Rudd et al., 2013; Neyland, 2017; Edmiston & Nicholls, 2018). Thus, stakeholders are often drawn to the potential of SIBs to open up a new source of capital to fund social change, particularly when public spending is constrained. In many instances, including in New York, agencies have identified several 'bold' ideas for social change, which exceed the total funding available, thus requiring private capital (Rudd et al., 2013). This driver is seen as particularly important for stakeholders such as smaller third-sector organizations that have experienced significant funding cuts; and also local governments who have faced budgetary pressures and are thus often searching for new ways to engage the private sector (Rudd et al., 2013; Hughes & Scherer, 2014; Cooper et al., 2016). For example, Greater London Authority (GLA) interviewees involved in the London Rough Sleeping SIB were committed to assisting the homeless. They saw the additional capital as a key benefit and motivation for pursuing the SIB.

Another regularly mentioned motivation was the ability of SIBs to transfer commissioning risk to service providers and investors (Tan et al., 2015; Rudd et al., 2013; Laing et al., 2017; Wallace & Roxas, 2014; Neyland, 2017; Tomkinson, 2012; Fraser et al., 2018; Cooper et al., 2016; Loxley, 2017; Ronicle et al., 2014; Oroxom et al., 2018; Ronicle et al., 2017). As the public sector, theoretically, only pays if the intervention is successful, SIBs lessen the risk that taxpayers will pay for ineffective programs and are thus often perceived as a less risky approach than a conventional contract (Ronicle et al., 2014). Despite this potential benefit, one study found that a quarter of the service providers not yet involved in a

SIB cited financial risk as a barrier to SIBs involvement (Ronicle et al., 2014). As risk transfer is often cited as a motivation for pursuing and benefit of SIBs, this was a surprising finding. This perhaps suggests that many providers are unaware of how SIBs can shift financial risk and/or are unable to take on risk in SIBs that require the provider to invest (Ronicle et al., 2014). This finding reinforces that SIB models and structures are appropriate for different stakeholders and pursued for different reasons.

Box 3: example (achieve greater funding certainty for NGOs)

> *"the St Mungo's staff we interviewed seemed to be dominated by the need to bring in revenue. The move to try out a radical new form of funding is understandable in view of budgetary challenges. David Evans, Finance Director for St Mungo's, is quoted as saying the potential for profit was "an added attraction when social protection funding is shrinking" (GreenFutures, 2013). Our Triodos interviewee also indicated that St Mungo's could earn as much as £300,000 in profit on the SIB"* (Cooper et al., 2014, p. 74).

3) Regulatory pillar

As highlighted earlier, the regulatory pillar refers to descriptions and judgements of rules and regulations related to SIBs by the government or other formal authorities (DiMaggio & Powell, 1983; Flickinger, 2009). While many jurisdictions have enacted SIB-focused regulations (e.g. the NSW Social Impact Investment Policy; and NSW Social Impact Investment Policy 2.0), which have contributed to the growth of SIBs, the included studies rarely referred to rules or regulations (Hawkins et al., 2017). Thus, no clear themes were identified for this pillar. This was an interesting finding, particularly when compared to the over-abundance of content related to the normative and pragmatic pillars. While no clear themes emerged, some general observations could be made regarding this pillar. For example, in some instances, regulations were mentioned in a manner that could discourage the use of SIBs; for example, one study noted that some investors (e.g. superannuation funds) may have avoided becoming involved in SIBs due to feeling unsure about whether they are

constrained due to regulations (KPMG, 2014). Some broader SIB studies that did not meet the inclusion criteria commented on the regulatory environment. For instance, Guter-Sandu (2018) noted that the institutionalization of social investment led not only to the launch of the initial Peterborough SIB but also to the enactment of the Public Services (Social Value) Act of March 2012, which required the principles of the blended value proposition to be observed by anybody commissioning public services.

4) Cognitive pillar

Cognitive legitimacy encompasses legitimacy based on comprehensibility and legitimacy based on taken-for-grantedness (Suchman, 1995). The SIB funding model is still highly contested, suggesting that 'taken-for-grantedness' has not yet occurred, and it has not become fully integrated into the institutional framework. Thus, legitimacy based on comprehensibility is the most relevant to SIBs; it relates to understanding the 'working principles' of the funding model (Aldrich & Fiol, 1994; Scott, 2013; Jansma et al., 2020, p. 197). As SIBs involve a change in governance, knowledge and understanding of the funding model are essential for gaining legitimacy, thus making the cognitive pillar important to consider (Aldrich & Fiol, 1994). Regardless, the cognitive pillar was far less prevalent than the normative and pragmatic pillars. Regarding motivations for adoption (or otherwise), some of the included studies discussed comprehensibility, for example, referring to what is known (or not known) about the effectiveness of SIBs, the risks associated with SIBs and the working principles. Many of the themes relevant to this pillar relate to the lack of knowledge on the 'SIB effect'; a lack of knowledge regarding crucial parts of SIB contracts, such as risk and return ratios, and a lack of measurement skills and knowledge among SIB stakeholders (Tan et al., 2015; Ronicle et al., 2016; DCLG, 2015; Neyland, 2018; Cooper et al., 2016; Loxley, 2017; Edmiston & Nicholls, 2018; Ronicle et al., 2014; Oroxom et al., 2018; Disley

et al., 2015). The three most commonly cited themes (evaluation/evidence challenges; harm/unintended consequences; and added complexity) are briefly discussed below.

The included studies frequently referred to evaluation challenges inherent in the SIB funding model. One challenge was the fact that some SIBs, such as the Essex Multi-Systemic Therapy (MST) SIB insist on *"payment through contractual obligation, regardless of the ability to attribute an effect"* (Neyland 2017, p. 499). In Essex, for example, investor returns were immediate as they could be made as soon as a child entered MST. This issue links to another frequently referenced issue; in a SIB, narrow outcome measures are typically used to facilitate fast repayment to investors (Loxley, 2017). The "success" or "failure" of a SIB is narrowly defined in terms of taxpayer savings; other social benefits may accrue but not be accounted for in the project's definition of success. In the policy area of OOHC, for example, the trigger measure is typically an immediate outcome (e.g. days of care averted) rather than a longer-term outcome such as the closure of residential care homes. This can become problematic if the short-term outcomes are not truly accurate proxies for the ultimate outcome and/or if they are not the most meaningful outcomes (Griffiths et al., 2016).

While included studies overwhelmingly focused on the benefits of SIBs, it is worth noting that many of them explicitly discussed potential harm/unintended consequences that could arise from the SIB funding model. For example, as outcomes-based contracts underpin SIB-funded projects, services are inevitably focused on achieving social outcomes rather than the continuity of service outputs that can characterize conventional public sector spending (Edmiston & Nicholls 2018). Some beneficiaries thus felt abandoned once outcomes had been achieved: *"I'm surprised by the way its just been ended so suddenly—- it's meant I'm not doing so well now"* (Edmiston & Nicholls, 2018, p. 70). Despite payment metrics being

constructed on the assumption that social outcomes will be lasting and thus result in significant cost savings, this concern raises questions over the extent to which social outcomes can reasonably be considered to have had a permanent effect on service users in a SIB contract (Edmiston & Nicholls, 2018). Additionally, it was noted that weak data could lead to unintended consequences and the misallocation of resources (Hughes & Scherer, 2014). Concerningly, it was highlighted that the SIB model could foster distrust amongst some beneficiaries who feel they are only being helped so that the provider can receive their payment.

Included studies commonly highlighted that SIBs are more complex than traditional funding mechanisms, with complicated partnership structures and processes required (Tan et al., 2015; Ronicle et al., 2016; KPMG, 2014; Rudd et al., 2013; DCLG, 2015; Rotheroe, 2014; Ecorys UK & ATQ Consultants, 2016a; Tomkinson, 2012; Hawkins & Bieretz, 2017; Warner, 2013; Cooper et al., 2016; Loxley, 2017; Ronicle et al., 2014; Van Es et al., 2016). For instance, *"the Ways to Wellness (WtW) SIB includes the Clinical Commissioning Group (CCG), a Special Purpose Vehicle (SPV), four service providers, the Big Lottery Fund, Cabinet Office and Bridges Ventures, leading to what WtW described as a "complex web of contracts"* (Ronicle et al., 2016, p. 18). In some SIBs, the presence of multiple commissioners (e.g. CCGs and a local authority) has presented additional administrative challenges, for instance, in obtaining sign-off from all actors on decisions. Many reports noted that the complexity of SIB contracts had caused hesitation among stakeholders. For example, in the NSW SBB Trial, SBB stakeholders *"recognized that social benefit bonds are one form of contracting or financing which can be more complex than other forms (such as a loan or an outcomes based funding model)"* (KPMG, 2014, p. 18). In one study, it was noted that *"a strong perception exists among some investors that the SIB model can be made*

unnecessarily complicated by both commissioners and intermediaries" (Ronicle et al., 2014, p. 48). The complex nature of SIBs has thus potentially deterred some investors, for instance, with an investor noting *that "we think there are easier and better ways to put our money to work"*; and *"it's hard to see commissioners wanting to talk about the structure of SPV's on a Monday morning"* (Ronicle et al., 2014, p. 46).

Box 4: example (lack of appropriate knowledge/skills in organizations)

> *"All interviewees rated their own individual understanding of SIBs as high, which might be expected as we were interviewing those with direct responsibility for SIBs and/or the evaluation of them as investment opportunities). However, interviewees indicated that the pool of individuals within their respective organizations that held a similarly high level of awareness and understanding of SIBs was generally narrow"* (Ronicle et al., 2014, p. 44).

6.3.5 Contribution of individual studies to the identification of themes

I also analyzed how each study contributed to constructing the final themes relating to possible motivations for SIB adoption (Appendix J). Some studies were over-represented and others were under-represented in the thematic framework. Regardless, sensitivity analysis (i.e. analyzing how the themes changed when certain studies were removed) suggested no clear association between factors such as study quality and groups of themes (Appendix L). Future studies could further explore associations between specific populations and themes.

6.3.6 Analytical themes

I identified three overarching analytical themes from the above analyses that describe the implications of the above findings, as discussed below.

1) *Instrumental reasons for adoption are regularly cited, despite limited evidence of SIB effectiveness*

While there is a dearth of quantitative evidence on SIBs, and thus, the effect of the funding model remains unclear, several included studies cited instrumental reasons as a main motivation for pursuing SIBs. In particular, many themes in the instrumental category focused on how SIBs are worthy of pursuing as they can improve outcomes, achieve better value for money and enhance accountability (e.g. Hughes & Scherer, 2014; Ronicle et al., 2014). This paradox is consistent with the public administration literature, which highlights policymakers' tendency to oversell a policy's benefits while downplaying or ignoring the disadvantages (Baier et al., 1986; Brunsson & Olsen., 1993; King & Crewe., 2014). Likewise, it is well-recognized that evaluations are often based on the analysis of rhetoric and ideology; robust measurement of the tangible costs and benefits is often absent (Hood & Dixon, 2015).

The management fashions literature gives additional context as to why instrumental reasons (e.g. better outcomes, increased efficiency etc) may be regularly cited in the qualitative literate, despite limited supporting (quantitative) evidence. Abrahamson (1996, p. 261) notes that management fashion setting is *"largely a cultural phenomenon, shaped by norms of rationality and progress".* In this context, *"organizational stakeholders expect managers to manage their organizations and employees rationally–- that is, by the most efficient means to important ends"* (Meyer & Rowan, 1977; Abrahamson, 1996, p. 261). In contexts which are characterized by high levels of uncertainty (e.g. in the policy area of SIBs), however, there may be ambiguity both in terms of what constitutes 'important ends' to pursue and also in terms of what the 'most efficient means' are to pursue those goals (March & Olsen, 1976). Under these circumstances, it is recognized that managers must *"create the*

appearance" that they are operating in accordance with norms of rationality (Abrahamson 1996, p. 261). Importantly, evidence may not necessarily be drawn upon or considered important, as perceptions can be sufficient for appearing to be progressive. Managers can be perceived as rational *"if they appear to use management techniques, such as quality circles, that stakeholders believe are efficient means to important end" (*Abrahamson 1996, p. 261). Rhetorics, rather than evidence, can create the belief that the techniques proponents champion are rational and at the forefront of management progress. Abrahamson (1996) holds that rhetorics vary in how they achieve this belief. Some management rhetorics indicate how a management technique constitutes an innovation; they reveal the history of successes and failures and disclose scientific evidence to show how the management technique represents an improvement, according to transparent criteria. Others *"provide no careful evidence that techniques constitute improvements. Moreover, they present old and forgotten management goals and innovations as if they had just been invented"* (Abrahamson 1996, p. 268). In the context of SIBs, it appears that the latter has occurred, with technical benefits regularly spoken about in the qualitative literature but not evidenced in the quantitative literature (Chapter 5).

2) *Particular pillars of legitimacy feature more regularly, although this provides an ambiguous signal regarding adoption motivations*

There was a noticeable difference in the extent to which different pillars of legitimacy were discussed, suggesting that some have played a more significant role in driving SIBs adoption. Themes regarding the normative pillar of legitimacy were easily identifiable, with extensive discussion about how and why SIBs are 'right' for society (Wallace & Roxas, 2014). These themes included value judgements and beliefs about how SIBs would be beneficial for society as a whole, for instance, by filling gaps in service provision, focusing

on particularly vulnerable groups and taking a long-term focus on outcomes (Mason et al., 2014; DCLG, 2015). Conversely, the pragmatic pillar (also frequently discussed) focused on the potential benefits to particular stakeholders. The themes identified in this category suggest particular power dynamics and vested interests, with many relating to the benefits for investors (e.g. obtaining profits, allowing for the recycling of capital, giving investors competitive advantage etc) (Hughes & Scherer, 2014). It was also noticeable that there was very limited discussion of the regulative pillar, even though SIBs have been progressed in some jurisdictions (e.g. NSW, Australia) largely due to very specific and proactive policies introduced by the government (KPMG, 2014). This suggests that SIBs have primarily been pursued for normative and pragmatic reasons, although some regulatory aspects have facilitated their usage in particular jurisdictions. The absence of the regulatory pillar was also evident when the pillars were coded for the random sample of tweets (see Chapter 8) as a robustness check; only 6.88% of the tweets featured the regulatory pillar. Consistent with the findings of this chapter, this was significantly less than the other pillars, including cognitive (36.30%), normative (46.42%) and pragmatic (35.90%), as well as instrumental motives (39.13%).

The cognitive pillar was also under-represented in terms of themes that can help to explain the adoption of SIBs; however, several themes were identified that could shed light on why SIBs have not been adopted in many instances and jurisdictions. For example, there was extensive commentary about the evaluation challenges surrounding SIBs (which have made it difficult to build understanding or generate evidence) and the need for more evidence regarding the effectiveness of the funding model (Disley & Rubin, 2014; Ronicle et al., 2016). The themes also emphasized that no clear blueprint exists for SIBs, findings are often not generalizable, there are a lack of appropriate knowledge and skills, and confusion exists

among stakeholders about many aspects of SIB contracts (Hawkins & Bieretz, 2017; Rotheroe, 2014). Thus, another finding is that while some of the legitimacy pillars mainly generated themes in support of SIBs adoption, others provided insights as to why SIBs uptake has been slower than expected and why some jurisdictions have failed to adopt the funding model altogether.

When considering what the findings reveal about particular pillars of legitimacy, it is essential to reflect on the types of studies that were included and the extent to which these studies included the opinions of SIB practitioners. While I initially searched academic databases, I also conducted an extensive grey literature search. Consequently, study types varied, with both academic and non-academic studies included. While still relevant, it must be noted that study types, such as academic studies, are one step removed from the motivations of practitioners. Thus, while they capture practitioner motives, these are mixed in with the selection and interpretive biases of the academics performing and reporting those studies.

3) *Building a unified theory from the pillars is difficult although they provide a useful framework for analysis*

While this thematic synthesis has highlighted that the various pillars of legitimacy are well-represented in the qualitative literature on SIBs (thus providing a useful framework), *"conceptualizing institutions as comprising these three disparate elements raises a number of problems"* (Greenwood et al., 2017, p. 729). For instance, each pillar is based on vastly different philosophical assumptions about the nature of social reality. This causes problems in creating a unified theory of institutions *"as it is difficult to see how to bridge the ontological differences between the underlying perspectives on which the pillars rest"* (Greenwood et al.,

2017, p. 729). Additionally, the pillars have very different dynamics, with both the regulative and normative aspects of institutions being *"the products of human design, and the outcomes of purposive action by instrumentally oriented individuals"* (DiMaggio & Powell, 1991, p. 8). Thus, for these pillars, stability and compliance is *"assured through values and norms that provide individuals and corporate actors consciously understood legal, social or moral rules and guidelines for behaviour"* (Greenwood et al., 2017, p. 729; Hirsch, 1997). As a result, there is an opportunity for actors to contest and deviate from institutions as they work to resist pressures that they have identified and perceive as misaligned with their interests (Hirsch, 1997). Conversely, the assumptions and mechanisms underpinning the regulative and normative pillars are rejected in a cognitive framework, as the cognitive pillar works with *"no questions asked and without deviation"* (Hirsch 1997, p. 1710).

Different inferences can be drawn from the prominence or absence of particular pillars. Greenwood et al (2017, p. 730) note that *"when something has become institutionalized no outside sanctions or controls are require"* as it has simply become 'taken-for-granted'. This is a notion that has been well supported in the institutionalist literature, for example, with Zucker (1991, p. 86) noting that *"direct social control through norms or sanctions (incentives or negative) is not necessary … applying sanctions to institutionalized acts may have the effect of deinstitutionalizing them … the act of sanctioning may indicate that there are other possible, attractive alternatives".* Thus, the notable absence of themes related to the regulatory pillar could point to the reform already being institutionalized (and thus control mechanisms not being required) (Greenwood et al., 2017; Zucker, 1991). However, this is an ambiguous finding as there appears to be an overabundance of themes related to the normative pillar, which has some similar assumptions and mechanisms to the regulatory pillar. Greenwood et al., (2017, p. 730) raise the essential

theoretical question of whether the regulative and normative pillars are *"important pressures in the process of institutionalization rather than elements of an institution"*. Overall, they note that there are limitations with applying the three pillars approach to understand what an institution is and how it comes to be. Thus, in Chapter 8 of this thesis, I will use a discursive perspective to overcome these limitations and better understand the social constructivist nature of institutionalization.

6.4 Discussion

In this chapter, I have outlined the findings from a thematic synthesis of a purposive sample of SIB-focused qualitative studies. This synthesis has identified the categories of factors (both instrumental and institutional) which have influenced the uptake of SIBs since 2010 in various public sectors. It has uncovered some clear themes relating to the well-documented legitimacy pillars, thus suggesting that institutional factors have played a critical role in the growth of SIBs. While this thesis hypothesizes that legitimacy is central to the diffusion of public sector innovation, it is essential to also consider competing drivers. Thus, instrumental themes were also considered and identified, although high levels of decoupling were evident regarding the instrumental themes (when comparing the findings of Chapter 5 and Chapter 6). The finding that both theories matter is not particularly insightful as it does not assist with understanding how the relative importance of these two drivers is dynamic and has shifted over time. Thus, it is also integral to consider the chronology of the qualitative research evidence.

Broadly, the timing of publication of included studies (including those in the broader sampling frame) shows a time lag in the literature, with few qualitative studies published in the early years of the diffusion of the idea. Out of the 181 studies included in the sampling

frame, only 20 were published between 2010 – 2014, thus meaning that there was little evidence, of any kind, in the initial years. A further 71 qualitative studies were published between 2015 – 2018; 89 were published from 2019 – present; and the date is unknown for one study. In chapter 5, I demonstrated that the quantitative studies that have used a robust design and reported positive, statistically significant effects have emerged more recently (e.g. ID Insight, 2018; Gallardo et al., 2021; Hunter et al., 2022), with varying levels of rigour used and mixed evidence reported, prior to that. Thus, SIB proponents have benefited from a prolonged period of uncertainty across both types of evidence, which has likely prevented the idea from being dismissed. The emergence of a growing body of rich qualitative literature from 2015 onwards, may have assisted proponents with continuing to propel the idea along, in the absence of robust quantitative data. Many studies failed to acknowledge their methodological limitations or the importance of using mixed methods and made strong claims about the benefits of SIBs.

More specifically, comparing the Chapter 5 findings to the analysis of qualitative evidence, over time (Table 6.8), showed that across all periods, a high level of decoupling exists surrounding the instrumental factors (and some institutional factors, but to a lesser extent) (Bromley & Powell, 2012). As described in Chapter 2 and Chapter 3, the term 'decoupling' was originally used to refer to a gap between policy and practice (e.g. policy remains unimplemented, evaluation not present or only symbolic). However, it was then expanded to also include a gap between means and ends (e.g. policy implemented but with an uncertain relationship to outcomes) (Bromley & Powell, 2012). An example of decoupling, identified in this chapter, is repeated references to governmental 'savings' (a claim lacking evidence in the literature) (Tan et al., 2015). Both types of decoupling are relevant for this context; for example, the gap regarding hypothesized and actual savings could exist due to

SIBs sometimes being launched but without the incorporation of outcomes that can lead to cashable savings (i.e. policy-practice gap); and/or due to SIBs being pursued despite little evidence of a known relationship between a SIB and the goal of generating cashable savings. Decoupling may help explain how and why the themes identified in this review are inconsistent with the effectiveness review findings (Chapter 5). The identified gap may provide evidence of institutional motivations, although it is an ambiguous signal which necessitates further investigation.

This review uncovered some important insights regarding the malleability of the SIB idea and the extent to which key contextual differences can account for the spread (or otherwise) of the idea. While many countries that have launched SIBs were not represented in the included studies, it was possible to gain some insights about the malleability of the SIB idea in different jurisdictions, given that seven countries featured, and many studies used comparative methods to compare the approach used in different countries. Also, the included studies discussed SIBs in both developed and developing contexts. Differences in factors such as the degree of government support make it particularly difficult to discuss SIBs in a de-contextualised way, thus underscoring the importance of comparative studies (many of which were included in this synthesis) that focus on a small number of jurisdictions. While the UK has launched more SIBs than the USA, the US programmes are much larger in scale than the Central Europe and UK SIBs (Hevenstone & von Bergen, 2020). In the UK, the implementation of SIBs has largely been policy driven (i.e., top down approach), with central government and international development funding (Outcomes Funds) being used to create a financial incentive for local commissioners. For example, the British Government leveraged money from the Big Lottery Fund to set up seven investment funds (totalling £191 million), which local commissioners could apply for to cover the SIB outcome payments (Carè & De

Lisa, 2019; Vecchi & Casalini, 2019). Yet the number of UK SIBs declined after 2018, when the Department of Work and Pensions' Innovation Fund expired, demonstrating the importance of Outcome Funds in scaling SIBs in the UK. Several of the included studies (e.g. Griffiths et al., 2016; ICF Consulting Services, 2019; Loveless, 2020) discussed Outcomes Funds such as the Innovation Fund, providing rich qualitative insights from stakeholders involved in SIB projects funded through a central Fund. In other countries (e.g Canada, US, European countries), more of a bottom-up approach has been used, with SIBs being promoted by private partners, mainly philanthropic investors such as foundations and charities (Vecchi & Casalini, 2019). This bottom-up approach was evident in many included studies that focused on a single program and jurisdiction (e.g. Galley et al., 2014; Hawkins et al., 2017; Hughes & Scherer, 2014). Given the malleability of the idea, and the key contextual differences that are evident across jurisdictions, it will be important for future studies to use a range of methods, including both systematic review methods that allow the author/s to go beyond the findings of individual papers (as this study has), as well as detailed comparative studies, which compare and contrast the use of SIBs in different contexts.

In reflecting on the findings of this chapter, it is integral to consider both the strengths and limitations. Regarding strengths, this qualitative systematic review is exhaustive and was conducted following PRISMA guidelines (Moher et al., 2009). It applied an institutional lens, which has rarely featured in the SIBs literature, particularly in studies using systematic review techniques. As this review included sensitivity analysis (Appendix L), it is possible to have some confidence in the findings, as the results did not change significantly when low-quality studies were excluded. Regardless, there are many limitations which will be overcome in the subsequent chapters of this thesis.

In particular, there are inherent limitations associated with ‘vote counting’ (Cwikel et al., 2000; Hedges & Olkin, 1980). The factors reported as motivations aligning to an instrumental and/or institutional perspective were initially aggregated to identify the most frequently cited motivations. This can be a useful initial step, although it can also be problematic as it does not consider factors such as study quality, study size or the weighting of each factor (Popay et al., 2006; Suri, 2000). It was not possible to compare the different motivations shared by different stakeholder types as included studies often did not clearly attribute perspectives to particular stakeholders. Also, given the relatively small proportion of the total evidence base included in the final sample, it was not possible to draw strong conclusions about how motivations have differed at different time points. This is a notable shortcoming as the diffusion literature makes it clear that motivations may change depending on when an innovation is adopted. For pragmatic reasons, the review only included papers published in English; however, there are various potentially relevant non-English publications, noting that SIBs have been launched by 38 countries since 2010 (GO Lab, 2023a). For practical reasons, the studies were not double-screened or double-coded. The majority of the included studies were not specifically focused on identifying motivations for adopting SIBs (although they all discussed this to some extent). Thus, I needed to use considerable judgement, to create codes and themes; the themes may have been different if they had been double-coded. While well-known sampling strategies were outlined and used, specific publication years and author types were over-represented in the sample of included studies, potentially biasing the results.

Further, the inclusion criteria focused on motivations for adopting SIBs as perceived by practitioners. However, it may be possible to identify factors affecting the uptake of SIBs, without relying on research participants' perceptions (e.g ethnographic studies). Thus, some

relevant evidence may have been overlooked in this synthesis. Using a consilience approach, the above shortfalls will be overcome in the next chapters.

In the mixed-method analysis of social media data – presented in the following two chapters of this thesis – I address many of the above limitations. By building a dataset spanning from 2010 – 2020 and analyzing the content and the networks and interactions, I aim to understand the social construction processes surrounding SIB innovation diffusion. In doing so, I also facilitate a methodological reflection on the limitations of past approaches (such as citation tracking), which have previously been used to understand the level of interest in an innovation over time.

6.5 Conclusion

In this chapter, I have considered what the qualitative evidence reveals about how legitimacy and efficiency drivers have influenced the uptake of SIBs, particularly over time, as levels of uncertainty (about intervention effectiveness) have changed. At the conclusion of the previous chapter, we knew that little evidence exists to support the effectiveness of SIBs. I observed that as time has progressed (and evidence has remained inconclusive), it seems that the likelihood of adopters pursuing SIBs for instrumental reasons has declined. We did not know what the qualitative evidence could reveal about other possible motivations. The thematic synthesis reported in this chapter was thus motivated by the observation that, while a rich body of qualitative evidence has been collected focused on SIBs (looking at factors such as motivations, promises and critiques), the literature has remained largely unsynthesized, and an institutionalist lens has rarely been applied to it. Applying a formal method of qualitative evidence synthesis to this body of evidence has uncovered a series of themes that have gone beyond the content of the individual studies. I identified an

overabundance of themes related to the normative and pragmatic pillars, with the cognitive pillar featured to a lesser extent and the regulatory pillar almost completely overlooked. These patterns were evident regardless of the publication year, providing further evidence of the mix of instrumental and institutional motives remaining constant over time.

Research synthesis experts do not unanimously agree that thematic synthesis is the most appropriate methodological approach; and large bodies of additional data exist that have not been incorporated into this review (Brennan et al., 2018). One particularly under-used type of data in this context is social media data, which is yet to be incorporated into an empirical SIBs-focused study. Thus, in the following chapters, I present a mixed-methods analysis of the available social media data to understand how the concept has been socially constructed to be perceived as a 'legitimate' reform, over time. This will build on the findings of this chapter, which have suggested that institutional drivers have played a key role in the adoption of SIBs, throughout the diffusion cycle.

Chapter 7: Sentiment Analysis and Network Analysis of Social Media Data

7.1 Introduction

The previous two chapters provided the findings of a quantitative and qualitative systematic review. The quantitative review (Chapter 5) identified a lack of quantitative data on SIB effectiveness, even more than a decade after the first reform was launched. This, it was argued, calls into doubt explanations of continuing SIB diffusion based purely on rational, "instrumental" adoption. The qualitative review (Chapter 6) built on the findings of Chapter 5 by identifying evidence, in the qualitative literature, of broader reasons for SIB adoption. Many of these reasons, for example, obtaining additional funding, only paying for outcomes, and filling a gap in service provision, can be linked to the concept of legitimacy and the institutionalist argument that the pursuit of legitimacy matters to public sector organizations as much, if not more, than instrumental performance. While Chapters 5 and 6 assisted with understanding the changing landscape of evidence over time, crucial gaps remain, which can only be overcome by exploiting the available social media data. Namely, the preceding chapters did not allow for a rich understanding of the tenor of public, practitioner, expert and consultant "chatter" about SIBs.

It is challenging to understand public narratives around SIBs from analyzing formal modes of discourse alone (i.e. academic papers, government reports, etc). In fact, relying on such materials alone could bias our understanding of the diffusion of any public policy innovation if we cannot access more informal discussions and narratives about the relevant idea. The advent of social media and the historical accounts of professional and political discussion that it enables provides a useful form of data which has, until now, not been used to examine the diffusion of SIBs. Recognizing this gap, this chapter, and the next, analyzes

the available social media data relating to the development of SIBs. The aim is to understand how and why this novel method of public service delivery has diffused across various public sectors since 2010. Specifically, by documenting both the incidence and content of SIB-focused tweets over time and between different countries and reform actors, this chapter maps the social construction processes that have surrounded the diffusion of the reform and contributed to its evolving 'legitimacy'. Overall, methodologically, I demonstrate that social media analysis represents a novel technique for tracing the diffusion of management fashions over time and between countries, for which previous studies have relied on academic citation counts, which are a poorly targeted and lagging indicator of organizational practice.

In this chapter, I statistically analyzed the entire dataset of Tweets to identify and visualize patterns in the data. This included both calculating descriptive statistics and conducting sentiment analysis to understand how opinions have differed over time. I compared these findings against other types of social media data (e.g., Reddit) and online data (e.g. Wikipedia Page Views, Google Trends and Google Ngram Viewer) to identify whether the trends are consistent across public forums and discussion groups. I then calculated various centrality measures to understand which stakeholders have been particularly influential in the Twitter network, thus playing a pivotal role in spreading the idea. As the analyses show, at all time periods, tweets containing a positive sentiment constituted a larger share of the dataset, as compared to negative sentiment tweets. Positive tweets were shared more widely than negative tweets, and tweets posted from countries that had adopted a SIB were far more positive than those posted from non-adopting countries. Additionally, stakeholders with direct SIBs involvement featured heavily in the breakdown of the most central (i.e. important) actors in the network. These findings support the notion that

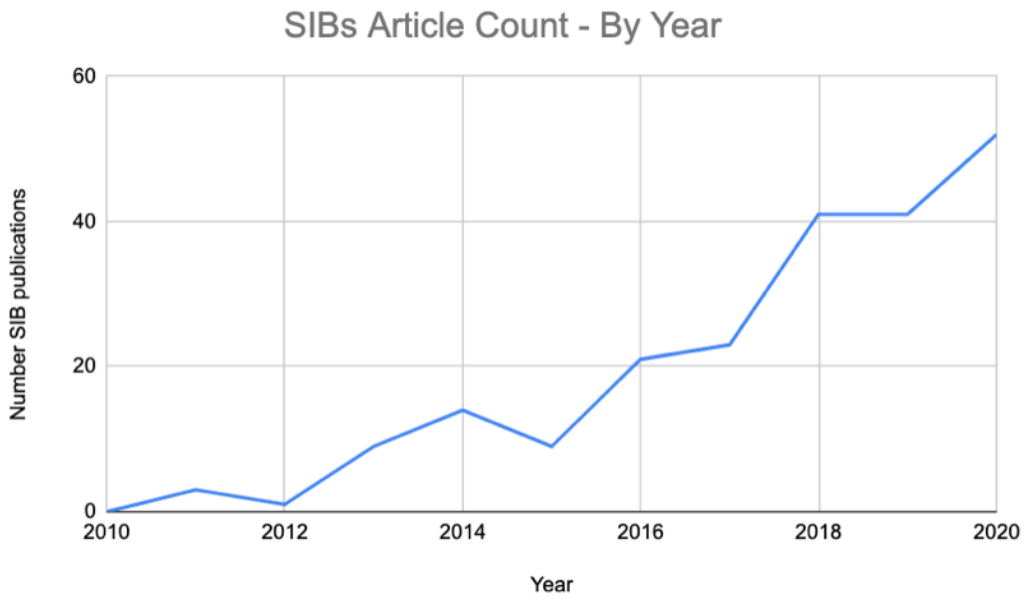
the growth of SIBs has been underpinned by institutional drivers, with factors such as novelty being overemphasized, while evidence has largely been overlooked in the online discourse.

7.2 Broad Trends Regarding the Development of the SIB Innovation

Before outlining this chapter's hypotheses, I discuss three key points that paved the way for the hypothesis development. Firstly, emerging forms of online data show a clear decline in interest over time (while academic citations suggest growing interest), thus demonstrating the importance of looking beyond academic evidence when considering how and why the reform idea has spread. Moreover, the contradictory data suggests that sentiments in the online "chatter" are different from those expressed in the academic literature and thus worthy of consideration. Secondly, actual SIBs usage has declined since 2016, suggesting that the decline in online interest is not simply due to taken-for-grantedness and may be indicative of the idea failing to become properly institutionalized. Thirdly, there has been a notable gap between interest in the idea and adoption, for instance, with a significant gap between the countries from which tweets have been posted and those where the reform has been adopted. While this is expected, to some extent, with all reforms, the gap is particularly stark in the case of SIBs, given how enthusiastically the idea has been promoted, across many years.

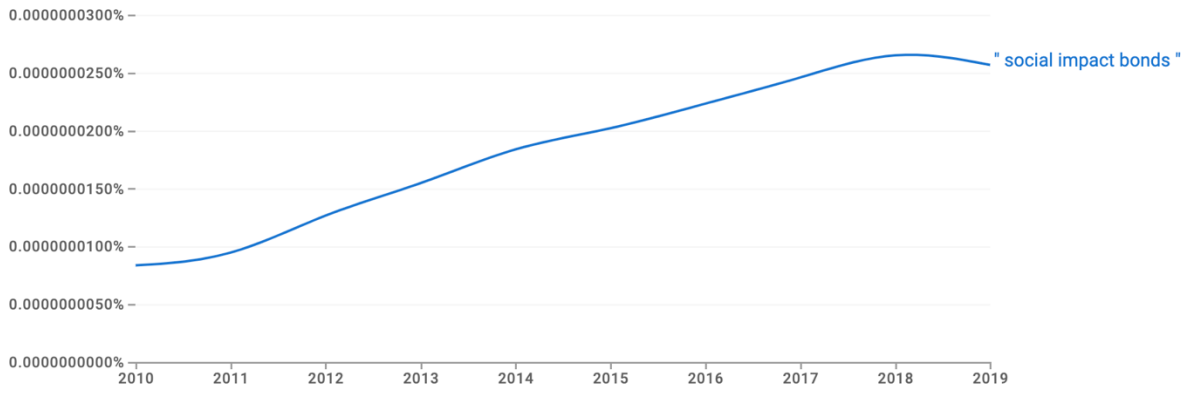
Figure 7.1 presents the citations count data for SIBs, showing that mentions of SIBs in academic publications have continued to increase over time.

Figure 7.1: SIBs Citation Tracking Data



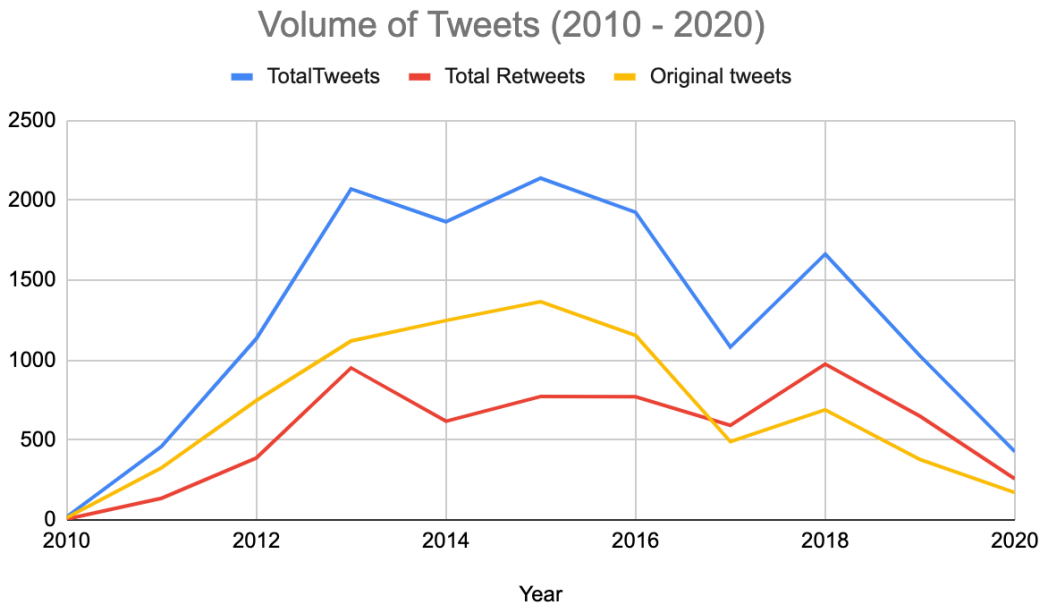
Similarly, Figure 7.2 presents data obtained using Google Books Ngram Viewer, which shows how the phrase “social impact bonds” has occurred in a corpus of books over time. While only a small percentage of the books have used the phrase “social impact bonds”, usage has steadily increased across time, only dipping slightly from 2018 onwards.

Figure 7.2: Google Books Books Ngram Viewer Data



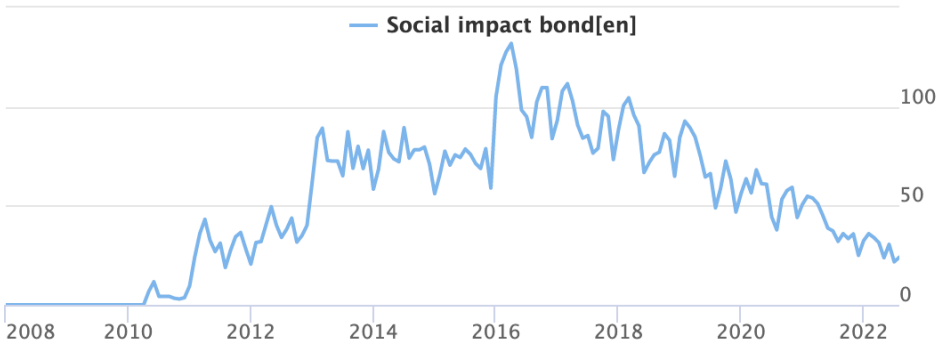
Regardless, other types of data suggest that SIBs have not necessarily become a lasting innovation and may instead represent yet 'another' management fashion. Declining interest over time could be identified when analyzing Twitter data, with interest largely dipping from 2015 onwards (Figure 7.3).

Figure 7.3: Twitter interest across time



Likewise, the declining interest could be observed when analyzing data from Wikipedia Page Views (see Figure 7.4). Wikipedia page views is a tool which allows users to instantly query Wikipedia pageview traffic data for a particular timeframe and keyword/s (Vardi et al., 2021).

Figure 7.4: Pageview Traffic Data for "Social Impact Bond (en)" Wikipedia Page

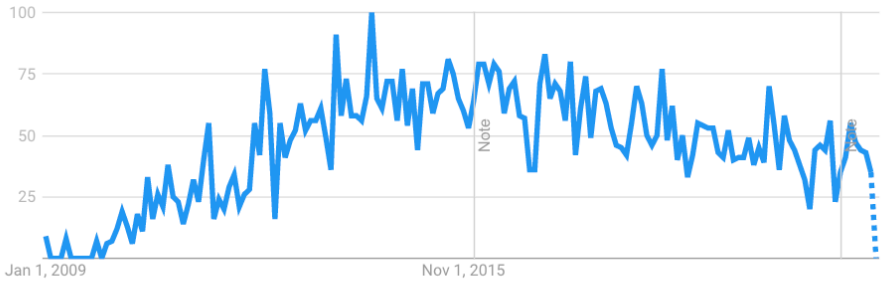


Source: Google Trends (https://www.google.com/trends)

Google Trends also provides valuable insights; this is a publicly available source of data which shows online Google search traffic. Users can visualize changes in time series data related to the public's online interest in particular keywords (Sato et al., 2021). The numbers in the graph below (Figure 7.5) *"represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term"* (Google Trends, 2022). As the line is trending downwards, it shows that the search term's relative popularity has decreased over time. It is also possible to view the locations where a search term was most popular during the specified timeframe, with values calculated on a scale from 0 to 100 (100 represents the location with the highest

popularity as a fraction of the total searches in that location). For this search term, the interest by region differed as follows: Australia (100); Netherlands (95); United Kingdom (89); Italy (73); Canada (57); United States (38); India (18); Chile (9).

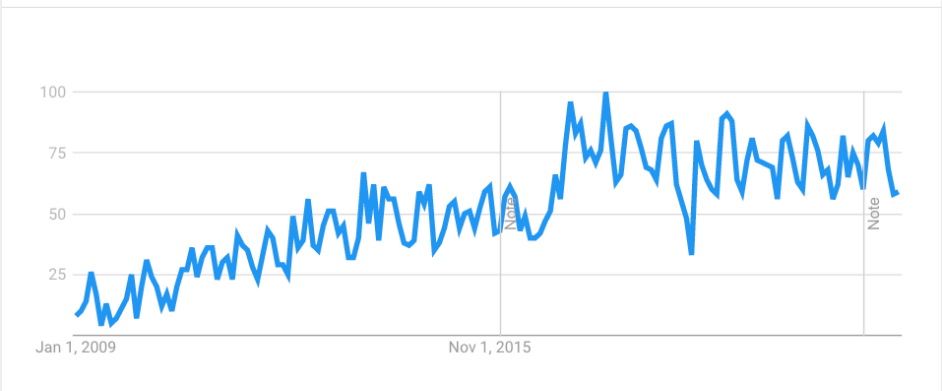
Figure 7.5: Google Trends Data for the Search Term "Social Impact Bond" (term search)



Source: Google Trends (https://www.google.com/trends)

Using Google Trends, it is crucial to search not only by "search term" but also by "topic", noting that topics are language agnostic and account for spelling variations and differing terminology. When searching by topic, the results differed (with far less of a decline over time), as depicted in Figure 7.6. The time series plot shows continual and consistent increases in the use of the term "social impact bond" until 2016, after which it becomes totally stable, and the upward trend disappears.

Figure 7.6: Google Trends Data for the Search Topic "Social Impact Bond" (topic search)



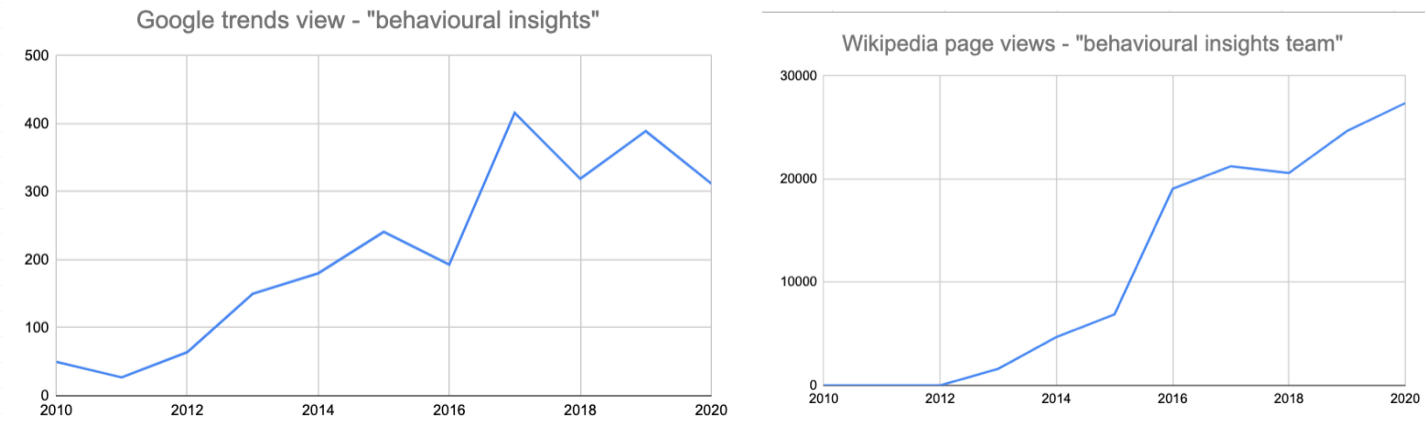
Source: Google Trends (https://www.google.com/trends)

When searching by topic, 33 regions were listed compared to only 8 countries for the search term. The top 10 regions include: Portugal (100); Japan (100); Brazil (91); New Zealand (88); Australia (87); Hong Kong (86); Singapore (80); United Kingdom (64); Philippines (61); and Canada (59). Notably, the related queries featured various non-English terms, thus accounting for non-English terms and variations.

When considered alongside the citation count data, the above patterns pose several interesting questions which require further consideration. For instance, it is unclear whether mentions of SIBs on social media platforms such as Twitter have declined because the idea has become institutionalized and is thus no longer viewed as 'novel', resulting in less discussion about the novelty of the idea. Tolbert and Zucker (1983, p. 25) suggested that one of the three indicators of institutionalization is that an idea is followed 'without debate', in addition to being widely followed and exhibiting permanence. Thus, the decline in Twitter mentions could point to increasing institutionalization, resulting in less debate and discussion about the idea; sentiment analysis will assist with further exploring if this is the case.

Conversely, the decline could reflect the fact that the idea is yet another example of a management fashion which has now become outdated and been replaced. Comparing the online interest in SIBs to another public sector reform that has been popularized ("behavioural insights") provided some further insights. On both Wikipedia and Google Trends, interest in behavioural insights has largely continued to increase (Figure 7.7). On a practical level, more than 200 institutions around the world are now applying behavioural insights to public policy.

Figure 7.7: Google Trends Views ("Behavioural Insights") & Wikipedia Page Views ("Behavioural Insights Team")



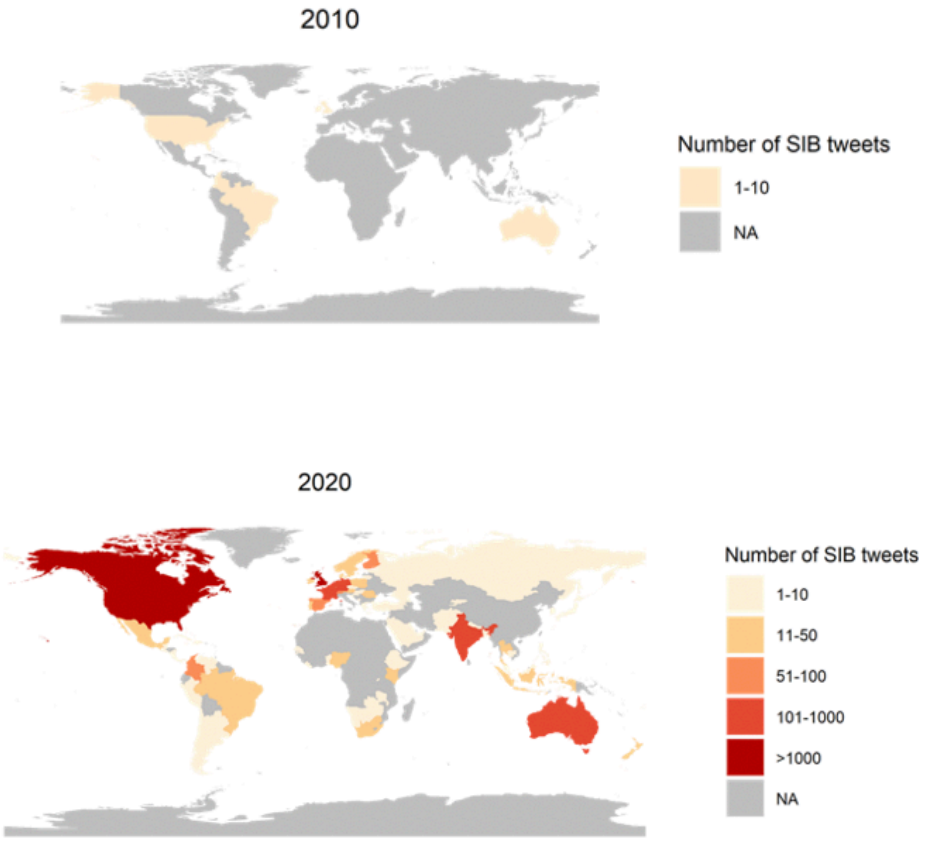
Source: Google Trends (https://www.google.com/trends)

Conceivably, the patterns identified in the SIBs social media data could also reflect more practical changes, such as: 1) a significant change in terminology (e.g. increased usage of terms like 'social outcomes contract' rather than 'social impact bond'); 2) a time-lag in terms of interest from the academic community; and/or 3) resourcing changes, for example, less resources being devoted to publicizing the idea online. While there are many possible explanations to be explored further in this chapter, the data on SIBs usage since 2010,

introduced in Chapter 1 (see Figure 1.2 and Figure 1.3) shows a continued tapering off, particularly since 2016. Along with decreased mentions in online forums, the idea has been adopted to a lesser extent over time, suggesting that the idea has not become institutionalized.

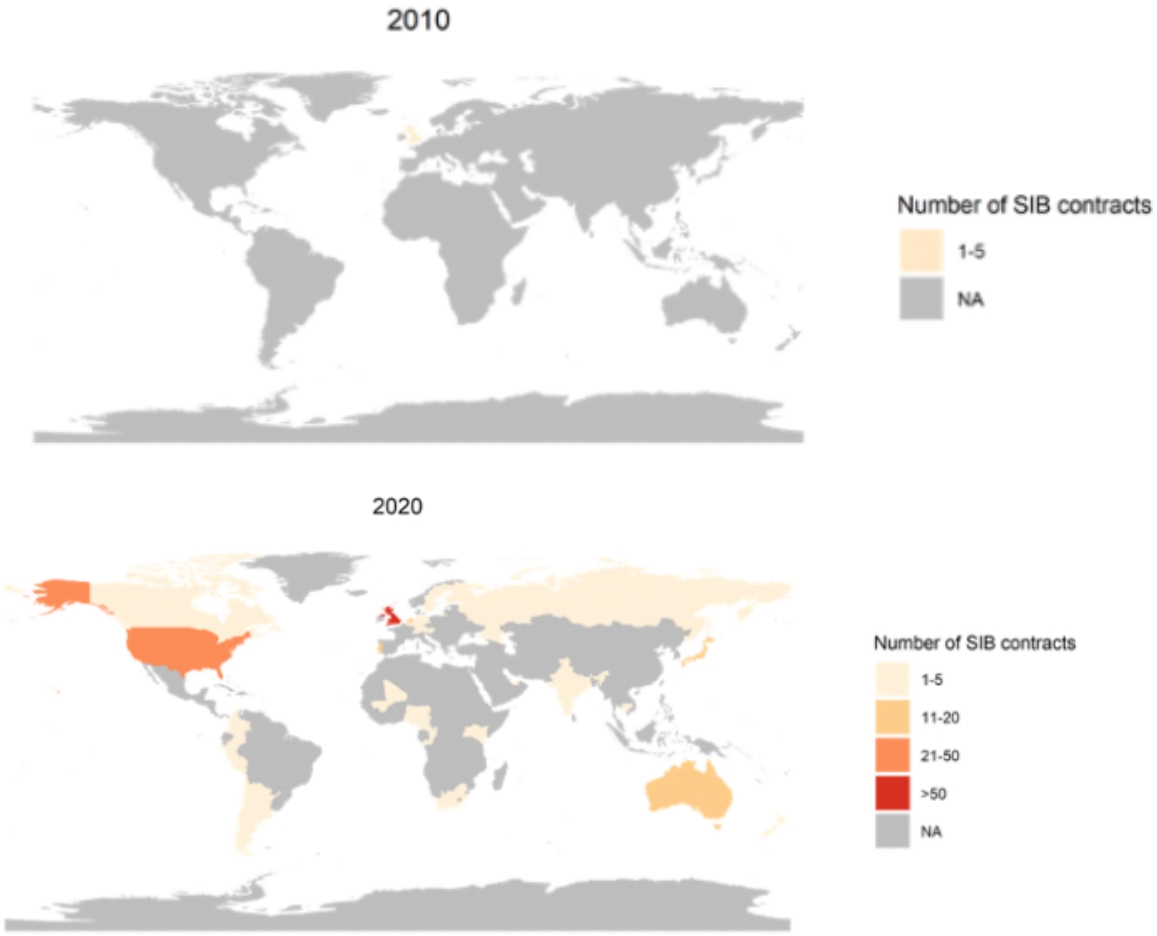
The location data obtained from the Twitter dataset also provides some interesting insights regarding the online spread of the idea. In an increasingly globalized world, the diffusion of social innovations transcends national borders; international organizations and other stakeholders influence settings at the global level with significant consequences nationally, particularly through online platforms such as Twitter (Schuster & Kolleck, 2020). Specifically, while a SIB is typically implemented at the domestic level, different stakeholders build networks and spread the idea and issues surrounding it globally, particularly through online forums such as Twitter (Torres Hernandez, 2008; Biermann, 2016). Figure 7.8 provides a visualization of the online interest (by location) in the idea over time, by providing a map view of the countries in which tweets had been posted from (up to and including the years of 2010 and 2020). Figure 7.9 compares this to the countries that had adopted at least one or more SIB in the same time periods.

Figure 7.8: Maps showing location of tweets containing #socialimpactbond mention over time



Source: Author analysis of Twitter data

Figure 7.9: Maps showing the spread of the SIB idea (i.e. launch of SIB) over time



Source: Author analysis of GO Lab (2022) data

Comparing figures 7.8 and 7.9 demonstrates that while the idea has continued to spread over time, many countries from which tweets have been posted have yet to adopt the idea in practice. Ultimately, the enthusiasm of policymakers and proponents has far exceeded and outpaced actual implementation (Roy et al., 2017).

The patterns identified above provide some interesting initial observations, particularly regarding declining interest over time, in the online "chatter", alongside growing academic interest. However, the above analysis did not consider the sentiments contained in the tweets or the extent to which particular stakeholders have influenced the online discussion. Thus, the remainder of this chapter focuses on filling these gaps. Sentiment analysis and network analysis are used to better understand the emotions and opinions conveyed in the tweets and to identify the stakeholders who have been influential in spreading the idea online.

7.3 Hypotheses Development

The data reported herein address two key research questions (outlined below), each related to an overarching hypothesis. The research questions and hypotheses have been developed to guide the analysis of the social media data and to overcome the limitations of applying past theoretical literature on innovation diffusion to the context of SIBs (discussed in Chapter 3 and Chapter 4).

R1: How have sentiments regarding SIBs changed over time on Twitter, and to what extent do the sentiments expressed in the tweets express emerging (inconclusive) research findings?

Related hypotheses: *as inconclusive (and sometimes negative) findings have accumulated in the academic literature, there has not been a similar growth in stakeholders tweeting negatively about the innovation, thus resulting in null and negative findings being overlooked in the online discourse.*

The instrumental viewpoint holds that administrative change is *"the result of deliberate goal-directed choices between alternative organizational forms"* (Brunsson & Olsen, 1993, p. 1); and that a 'chain of cause and effect' runs from the intentions of reformers to new structures and progresses, to changes in behaviour and improved results. In the case of SIBs, however, there is limited evidence of effectiveness, and the idea has spread regardless of whether it 'works' or not. Thus, it is hypothesized that, in alignment with the institutionalist literature (e.g. DiMaggio & Powell, 1983; Tolbert & Zucker, 1983), adoption decisions have often been driven by broader factors such as a desire to conform to prevailing (socially constructed) norms, with a lack of congruence expected between the sentiments expressed in the academic literature and those expressed in the online discourse. Notably, it is expected that positive tweets have been shared more widely than negative tweets and that online forums such as Twitter have effectively amplified positive sentiments about SIBs while overlooking any mentions of unintended consequences and harm.

R2: Which stakeholders have played a central role in communicating and promoting the SIB idea in online forums such as Twitter?

Related hypotheses: *Organizations within the "fashion setting community" have promoted SIBs in the Twitter discourse; thus, SIBs have diffused to many public sectors when organizations in fashion setting networks have promoted them, whether or not there is evidence of success.*

As noted earlier, in addition to contrasting the instrumental and institutional viewpoints (see section 3.4), some additional possible explanations will be explored in this thesis. It is well understood that fads or fashions can facilitate the diffusion of technically inefficient administrative technologies, thus providing another, related theory as to why SIBs have continued to diffuse to many public sectors despite limited evidence of effectiveness (Abrahamson, 1991). Under this lens, fads or fashions can fulfil a symbolic function, such as signalling innovativeness, without boosting performance (Abrahamson, 1986; Fombrun, 1987). Various researchers have noted that our culture's fascination with change has contributed to the ongoing search for new fads and fashions, with advantages often stemming from being aligned with growing fashions (Best, 2006: Spann, 1981). Thus, this hypothesis assumes that a powerful set of actors (such as consulting firms and investing firms) have propelled the diffusion of the idea, by promoting it within their fashion setting networks. Many of these actors have selected the SIB innovation, knowing that they can market it profitably, therefore impelling the diffusion of SIBs regardless of whether it is technically efficient (and possibly contributing to the rejection of more efficient innovations).

7.4 Overview of Methods

Social media data, which is produced in real-time and thus more immediate than other forms of data, which can be analyzed qualitatively as well as quantitatively, and which is authored by practitioners as well as academics and other observers, is used in this chapter to understand SIB diffusion. Unfortunately, particularly in the aftermath of the Cambridge Analytica data breach, many social media platforms have limited the data provided through their APIs. Nevertheless, Twitter continues to provide its data to the academic community (Ngai et al., 2015). Importantly, Twitter data are easily searchable using keywords as hashtags, denoted by the symbol #. Thus, I built and analyzed a dataset of all Tweets posted

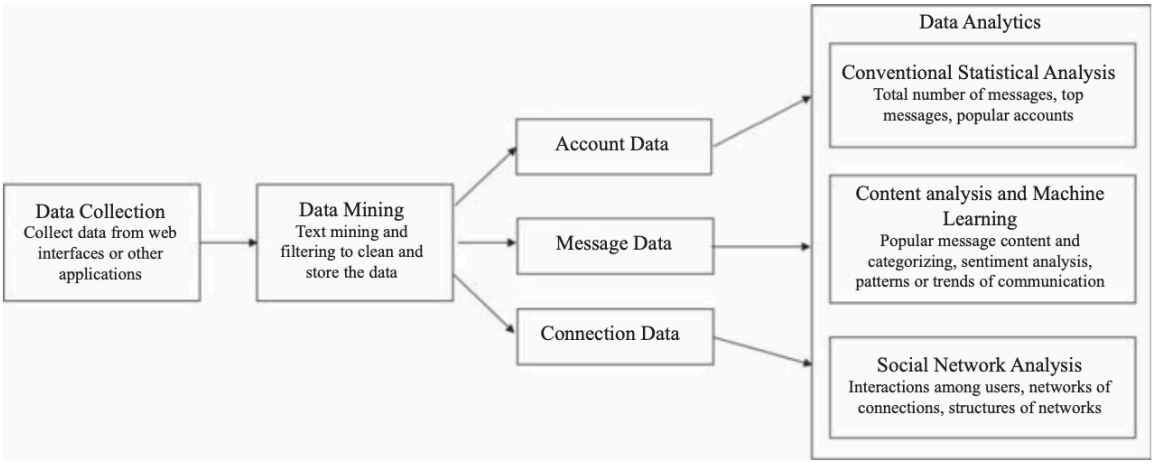
between 2010 – 2020 that used a relevant hashtag (#socialimpact bond; #socialimpactbonds; #socialbenefitbond; #socialbenefitbonds).

Hashtags involving other relevant acronyms (e.g. SIB, PbR, OBC etc.) were tested but found to be too broad; in many instances, the relevant acronym was used for a completely unrelated concept or context. Other related terms (e.g. #paymentbyresults, #outcomesbasedcommissioning, #resultsbasedfinancing) were also tested but excluded from the final dataset. The term #paymentbyresults also retrieved tweets that were too broad in scope; the tweets discussed all types of PbR (e.g. NHS funding) rather than only SIBs. Similarly, #resultsbasedfinancing retrieved tweets about non-SIB projects, as exemplified by the fact that projects launched in non-adopter countries were mentioned (e.g. Haiti). Additionally, #developmentimpactbonds was tested, but the distribution and descriptive statistics were found to be different for this dataset versus the SIB-specific one; thus, it was determined to be more appropriate to keep them separate (particularly since DIBs represent a clear domain shift). There are some additional terms of relevance (e.g. 'social impact finance'); however, it was practically challenging to include broad terms where the hashtag is used to discuss a variety of funding mechanisms, as a significant number of non-SIB-focused tweets were retrieved, thus diluting the dataset.

The final dataset included 13,836 tweets (7,703 original tweets and 6,113 retweets) collected using Twitter's Full Archive API. Additional data were gathered and analyzed from Reddit (a social news aggregation, web content rating, and discussion website), although the primary focus of this chapter was on Twitter data. Crucially, the dataset included a mix of stakeholder types (both SIB adopters and non-adopters), allowing for rich insights into how and why the innovation has spread. Diverse techniques are available to analyze data collected

from SNSs such as Twitter, including descriptive statistical analysis, content analysis, sentiment analysis, and social network analysis (Hu, 2019). I followed the three-fold approach advocated by Hu (2019) and depicted in Figure 7.10.

Figure 7.10: Social Media Data Analysis Framework



Source: Hu (2019)

Descriptive statistics:

The data were analyzed quantitatively in R using relevant statistical packages for this component. The descriptive statistics were calculated using the entire dataset to ensure that all of the available data were drawn upon in the study. The following were considered:

1. Tweet volume (all tweets, retweets, original tweets) by year, to understand how the popularity of the idea has changed over time;
2. The geographic spread of tweets over time, to understand the locations in which the idea has been discussed (and to identify any mimetic behaviour, such as neighbouring countries mimicking each other); and
3. The most frequent tweeters, and most frequent re-tweeters, by year, to understand who has been particularly active in promoting the idea.

I also linked the Twitter data with other data sources, for instance, through:

1. Comparing the volume of tweets to the volume of academic papers over time to identify the extent to which the academic evidence base has lagged behind the online "chatter" about SIBs;
2. Identifying which papers from the systematic reviews (Chapter 5 and Chapter 6) have been explicitly mentioned in the tweets to see the extent to which evidence has been used and referred to in the online discourse; and
3. Comparing countries where tweets have been posted from versus countries where SIBs have been adopted, using the GO Lab's International Network for Data on Impact and Government Outcomes (INDIGO) database to identify links between awareness (tweet posted from relevant country used as a proxy) and SIB uptake (i.e. launch of one or more SIB projects).

Sentiment analysis

Sentiment analysis integrates statistics, natural language processing, and computational linguistics; it detects opinions in natural language texts (Pang & Lee, 2008; Misuraca et al., 2020). It is crucial in the diffusion of innovations as ideas spread based on a social process influenced by sentiments. One of the main tasks involved in sentiment analysis is sentiment classification, in which texts are categorized according to the polarity of opinion (i.e. positive, neutral or negative opinion) (Ravi & Ravi, 2015). Lexicon-based approaches use a pre-existing dictionary; sentiment values from the dictionary are assigned to all positive and negative words or phrases. A combining function (sum or average) is then applied to make a final prediction regarding the overall sentiment. This approach avoids the need to generate a labelled training set. Conversely, machine learning approaches involve using a labelled training set to build a classification model which can predict the class of document based on pre-defined categories.

In this thesis, multiple pre-trained sentiment models (available in R) were used and compared to understand how sentiments contained in the online "chatter" about SIBs have evolved across time. Table 7.1 outlines the various models that were used. The pre-trained

models calculated a sentence-level sentiment score for each tweet; and classified each tweet into a sentiment category (positive, neutral or negative).

Table 7.1: R packages Used for Sentiment Analysis

Package	Author(s)	Launch date	Number of downloads	Custom dictionary	Weighted dictionary	Multilingual
syuzhet	M. Jockers	2015-02	343,840	Yes	Yes	Yes
RSentiment	S. Bose	2016-05	35,446	Yes	No	No
sentimentr	T. Rinker	2016-08	72,393	Yes	Yes	No
SentimentAnalysis	S. Feuerriegel & N. Pröllochs	2017-06	33,598	Yes	Yes	No
meanr	D. Schmidt	2017-06	7,907	No	No	No

Source: Misuraca et al., (2020)

Social network analysis

Rogers (1995) notes that decisions about adopting an innovation depend highly on communication with others. Communication networks emerge from exchanges between actors about ideas and experiences; particular stakeholders act as 'policy entrepreneurs' and thus play a crucial role in spreading ideas (Mintrom & Vergari, 1998). Evidently, *"we must understand the nature of networks in order to understand the diffusion process"* (Rogers, 2003, p. 331). Within this, it is vital to understand which actors use which communication channels; and which networks are formed as a result. Table 7.2 outlines the well-accepted measures of centrality that will be presented in this section of the thesis (Rochat, 2014).

As the Twitter dataset contains valuable user account information and information on retweets and interactions, social network analysis will also be conducted to understand how the reform community is organized (Wasserman & Faust, 1994; Borgatti et al., 2009). This will assist with identifying the direct and indirect interactions among users (Borgatti et al.,

2013). The results of this section will assist with identifying which actors have held central positions in the network, enabling them to exert significant influence on the diffusion of the idea.

Table 7.2: Summary of Key Centrality Measures to be Considered

Measure	*Why important to measure*	*What high score shows*
In-degree	Captures the community's level of engagement with the relevant user.	Users with a high in-degree score gain attention from their tweets among the relevant community; i.e. they are 'conversational hubs'. The information they initiate cascades throughout the network. As in-degree refers to the total number of retweets, mentions or replies, it can indicate prestige or popularity.
Out-degree	Captures the level of outreach of a user to the community.	This score measures the extent of a Twitter user's activity. Users with a high out-degree score engage highly with network members; they tweet a lot about a particular topic, thus aiming to reach the attention of other users by mentioning and/or replying to them.
Betweenness	Identifies users in a 'bridge' position between subdivisions of the relevant network.	Users with a high betweenness score are highly connected; they typically have the shortest path between the vertices of the graph (i.e. as more people depend on a user to make connections, their betweenness centrality score becomes higher).
Eigenvector	Recognizes that connections to more influential vertices are more critical than those with less influence.	Users with a high eigenvector score are particularly connected within the larger group; the score is assigned based on neighbouring scores.

7.5 Findings

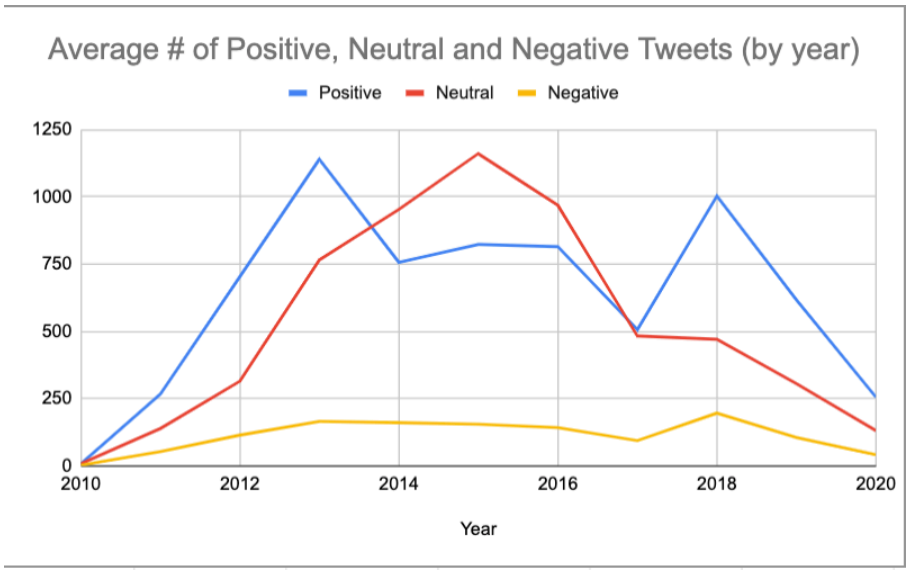
R1: How have sentiments regarding SIBs changed over time on Twitter, and to what extent do the sentiments expressed in the tweets reflect those expressed in the academic literature?

The accessibility and availability of the internet have enabled people to easily share their opinions online, particularly through forums such as Twitter (Jindal & Aron, 2021). Sentiment analysis allows for emotions to be analyzed, recognized and categorized; this is essential to diffusion studies, as innovation diffusion is a social process shaped by the changing sentiments of people over time (Rogers, 1995). In the field of natural language processing, the terms 'sentiment' and 'emotion' are common. Jindal and Aron (2021, p. 2) specify how emotions and sentiments differ:

> *"An emotion is a precognisant articulation of an individual affected by social and cultural outlooks. However, the sentiment is a conscious disposition of an individual based on his feelings and is persistent. The psychological and cognitive aspects of both these terms are considered as two different notions. Sentiments are a step ahead than emotions because psychological extents do not restrain them".*

This study classified tweets into three sentiment levels: positive, neutral, and negative. Various sentiment analysis packages were applied in R to identify patterns (and the results from the packages were compared to results obtained from hand coding, in Figures 7.12 – 7.15). The packages used different pre-trained sentiment libraries. Figure 7.13 shows the average number of tweets identified for each category per year. The figures were calculated by averaging the results from the five pre-trained models to give a pooled score. Figure 7.11 demonstrates that the dataset overwhelmingly features positive and neutral tweets, with negative tweets being noticeably absent throughout the entire time period.

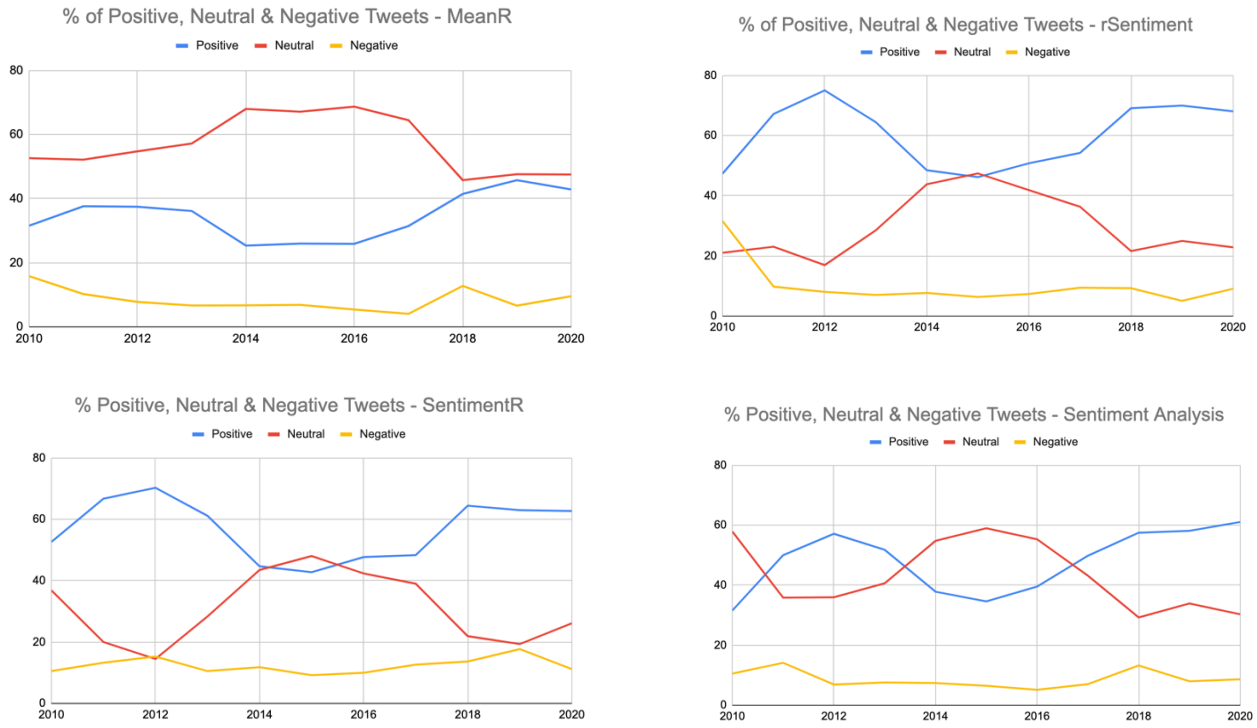
Figure 7.11: Average Number of Tweets Containing Each Sentiment Type

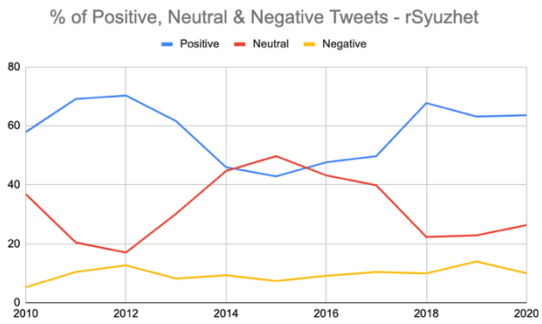


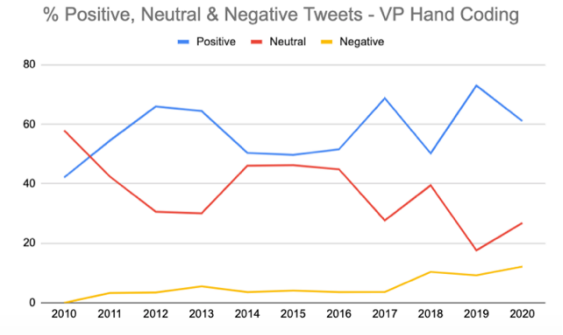
Each package identified a slightly different volume and percentage of positive, neutral and negative sentiments (Figure 7.12). It is noticeable that the MeanR sentiment package often identified fewer negative sentiments and labelled more as neutral. The existing Twitter

literature provides some insights into why this may have occurred. For instance, of the various packages used, meanR is the only one which does not provide the option of calculating polarity scores with a customized dictionary. Thus, it is impossible to tailor the dictionary according to research needs or apply a different weighting scheme to calculate scores (Misuraca et al., 2020). Existing studies have identified it as the fastest package for calculating polarity scores but the least accurate (Misuraca et al., 2020). The inaccuracy of MeanR was also consistent with the hand coding results, as I identified the lowest percentage of matching sentiments, when comparing my hand coding results to the sentiments assigned by MeanR (51.97%). The highest percentage of matching sentiments was found with Rsyuzhet (56.98%), followed closely by Sentiment R (56.74%), Sentiment Analysis (56.63%) and RSentiment (55.96%).

Figure 7.12: Percentage of Positive, Neutral & Negative Sentiments – for Each Package







Regardless of the differing scores, the overarching patterns were strikingly similar, thus demonstrating that the results are robust to various ways of measuring sentiment, including both automated and manual coding (Figure 7.13, Figure 7.14, Figure 7.15).

Figure 7.13: Visual Analysis of the Changing Sentiments Over Time (Positive Sentiments)

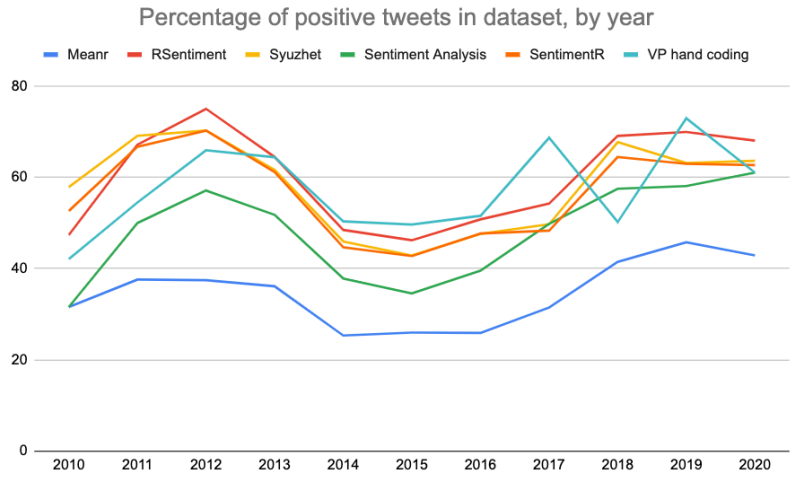


Figure 7.14: Visual Analysis of the Changing Sentiments Over Time (Neutral Sentiments)

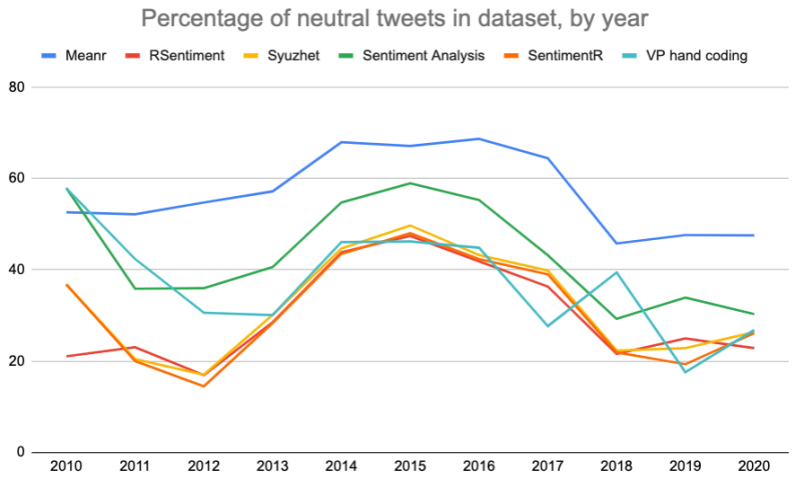
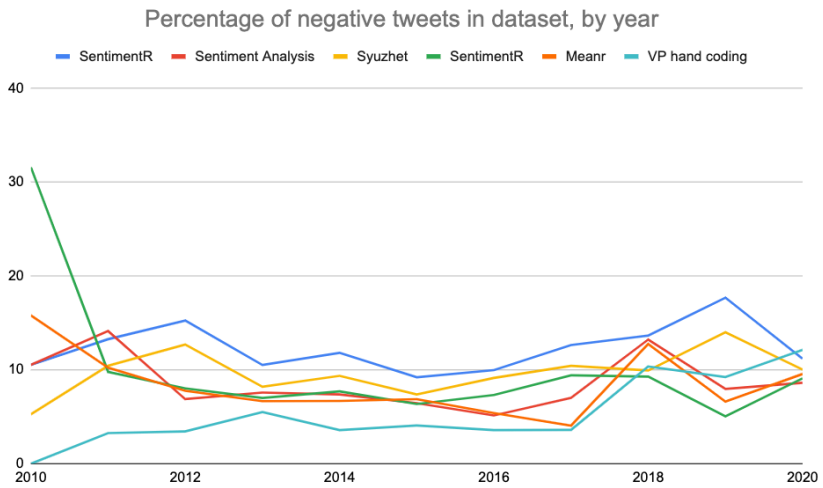


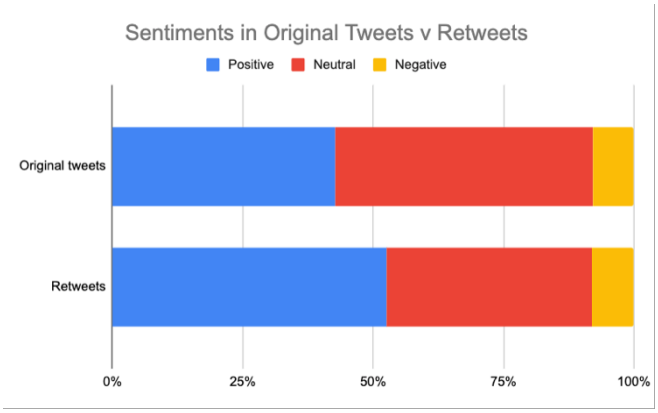
Figure 7.15: Visual Analysis of the Changing Sentiments Over Time (Negative Sentiments)



The visualizations above demonstrate that tweets containing negative sentiments (often discussing harm, failures, unintended consequences etc) have not featured highly in the Twitter dataset at any time point. However, positive sentiments have differed over time. They appeared to increase between 2010-2012, likely in alignment with the excitement and hype surrounding the launch of the initial Peterborough SIB (Disley & Rubin, 2014). Positive sentiments then dipped between 2012 - 2016, likely as initial enthusiasm declined and some failures (e.g. Riker's Island) emerged in the media and literature (Parsons et al., 2016). Interestingly, positive sentiments appear to have increased again from 2017 - 2020, perhaps related to the following three trends: 1) the emergence of some positive findings, such as the successful Educate Girls DIB, which was measured robustly and delivered strong investor returns and social outcomes (IDinsight, 2018); 2) growing interest in DIBs (which switched the innovation to a new domain) and the possible application to developing settings to overcome funding gaps associated with the Sustainable Development Goals (SDGs) (Alenda-Demoutiez, 2020); and 3) interest in the possible use of SIBs in the COVID recovery, for example, to obtain upfront funding for vaccines (Mair, 2021).

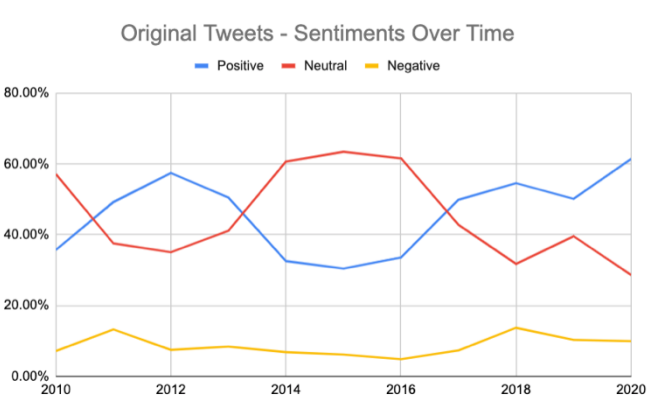
Some further insights were obtained by comparing the sentiments in original tweets to those in retweets. Negative sentiments were noticeably absent in both original tweets (7.86%) and retweets (7.98%), although the retweets contained slightly more positive tweets (52.61% v 42.65%) and fewer neutral tweets (39.41% v 49.49%) (Figure 7.16).

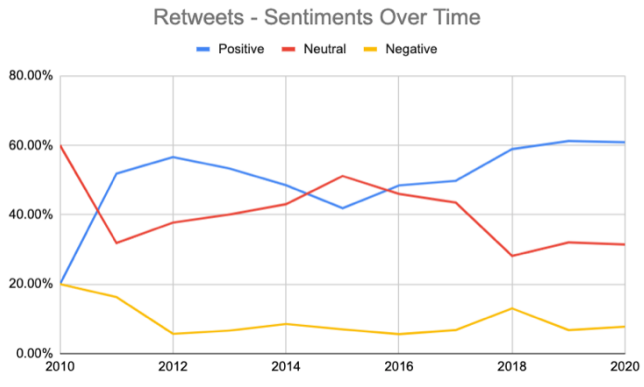
Figure 7.16: Sentiments Contained in Original Tweets v Retweets



The split between tweets and retweets over time revealed slightly different patterns. Broadly, the proportion of negative tweets remained particularly low in both retweets and original tweets; however, the split between positive and neutral tweets differed slightly (Figure 7.17). For the retweets, the positive tweets typically outweighed neutral tweets, although the split between positives and neutrals was more even in the original tweets.

Figure 7.17: Comparison of Sentiments Contained in Original Tweets v Retweets – Over Time

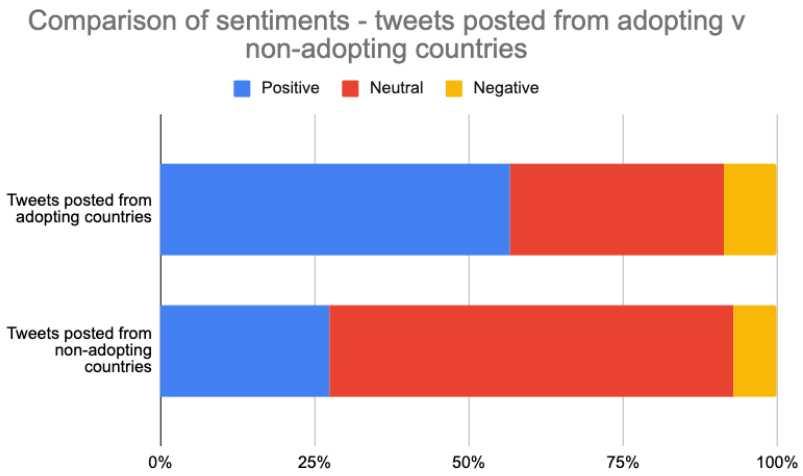




By comparing the sentiments in tweets posted from adopting countries versus non-adopting countries (Figure 7.18), it was possible to gain some insight into how sentiments have differed by the degree of direct experience. Of course, this is not a perfect proxy for the level of SIBs involvement, as an individual tweeter may not be directly involved with a SIB or influenced by whether their jurisdiction has launched a SIB or not. However, given the size of the dataset, it was infeasible to code every tweet with the tweeter's level of SIBs involvement. Thus, using the location data combined with SIBs specific data (countries that have adopted a SIB) was the most practically feasible approach. Figure 7.20 shows that tweets posted from countries that had adopted a SIB were far more positive than those posted from non-adopting countries (56.79% vs 27.48%).

Additionally, far more tweets with a neutral sentiment were posted from non-adopting countries (65.46% vs 34.53%). Thus, while the proportion of tweets containing a negative sentiment was similar, tweeters from the non-adopting countries were less optimistic and more neutral in their opinions towards SIBs. This further suggests (although in an imprecise way which requires further investigation) that the positive sentiments are mainly being pushed forward by SIB stakeholders with direct involvement (likely including both consultants and those implementing the SIB), many of whom directly benefit from the success of the SIB concept.

Figure 7.18: Comparison of Sentiments Among Adopter v Non-Adopter Countries



A further component of the analysis involved considering whether the positive tweets from adopting countries were posted at (or prior to) SIB launch versus at the time of SIB evaluation. This indicates whether the tweeter is positive and optimistic (i.e. no experience yet but enthusiastic) or positive based on true experience of results. In calculating this, it is crucial to account for the 'honeymoon' period, which can cause a lag effect whereby initial negative feedback may be protected. If a new relationship commences with an 'initial stock of assets' (such as favorable prior beliefs, trust, goodwill, financial resources or psychological commitment), *"the risk of the relationship dissolving at its inception is reduced, even if the initial outcomes of the relationship are unfavorable"* (Fichman & Levinthal, 1991, pp. 434-444). The duration of the 'honeymoon' period varies with the magnitude of the initial assets. To account for this, the analysis below (Figure 7.19, right graph) includes a 'pre + 2 years" category. Figure 7.19 demonstrates that the positive sentiments increased post-launch, both with and without the 'honeymoon' period accounted for. This provides some contradictory data to the above sections of this chapter, as it suggests that the sentiments may have been linked to results.

Figure 7.19: Comparison of Sentiments – Pre v Post Adoption in Adopting Countries

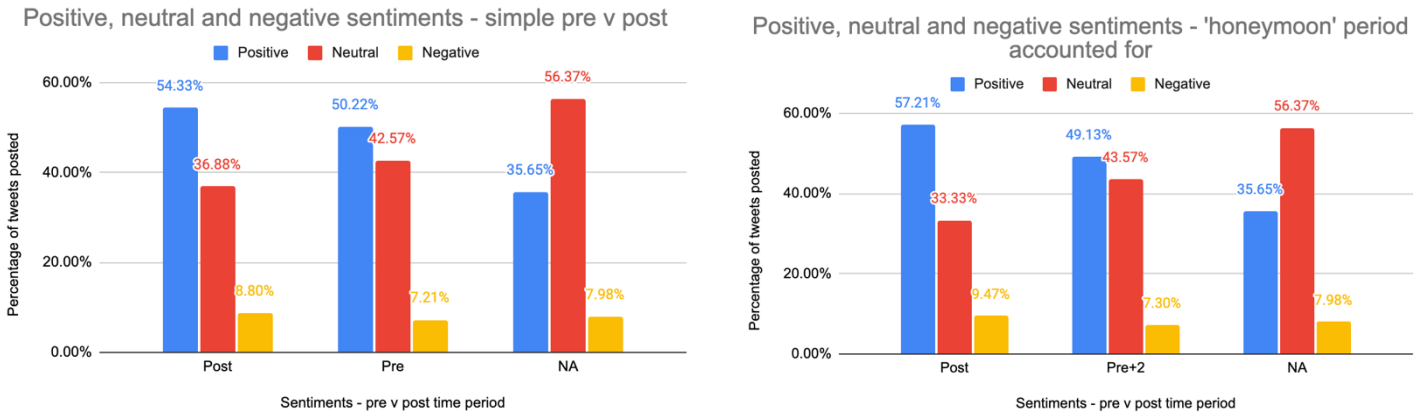


Table 7.3 contains sample tweets representing different sentiment levels based on manual annotation, and Table 7.4 contains (accurate) sample tweets for different sentiment levels based on the pre-trained models that were applied to the entire dataset. In some instances, the relevant tweet was coded as a particular sentiment by all five pre-trained models. For instance, each model identified the following tweet as positive: *"One year on, the #Newpin #SocialBenefitBond has helped many families and generated positive returns to investors - http://bit.ly/newpin-sbb"*. In other instances, the tweets were coded differently by different packages. Finally, in some cases, the accurate sentiment was only identified through email interactions with the author, which revealed nuances that had been overlooked. Thus, even when manually coded by a human, there is still significant room for error and misinterpretation, given the short length of a Tweet and the extent to which meanings can be implied rather than explicit.

Table 7.3: Sample Tweets Representing Different Sentiment Levels - From Manual Coding

Level	Sample tweet
Positive	*"New weapon available for governments #SocialImpactBonds cut costs & get #value4money pay for results only #DisruptAid"*
Neutral	*"Food for thought - Key note speaker @cabinetoffice informed us on #socialimpactbonds today"*

Negative	*"Inside trading, investing in revolving door insider climate change ideas or humanitarian aid, profiting from war, & disaster #SocialImpactBond create an emergency which gets your friends contracts 🤪usa 🇺🇸"*

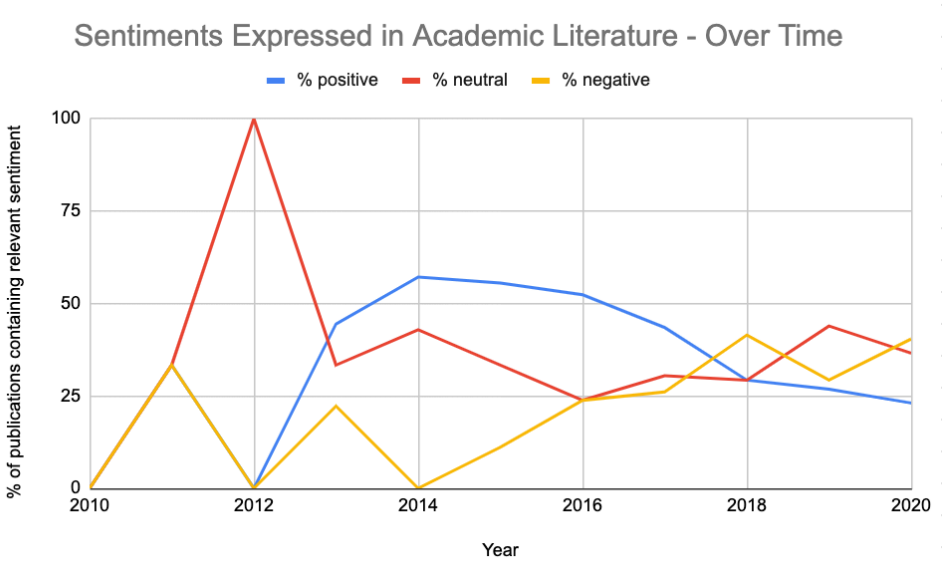
Table 7.4: Sample Tweets Representing Different Sentiment Levels - Based on Pretrained Models (examples of accurate coding)

Level	RSentiment	Sentiment Analysis	Syuzhet	SentimentR	MeanR
Positive	*"#obama is hoping to increase oversight and evaluation in the nonprofit sector with #socialimpactbonds http://ow.ly/3Ti4T thanks #nyt"*	*"@writerpollock agree! I'm so excited he proposed it, the pilots have been quite successful in the UK. Should be fun! #socialimpactbonds"*	*"@robintuddenham Huge potential for #socialimpactbond, would be interested to hear about any developments..."*	*"NYT - #SocialImpactBonds could improve gov't performance; benefit #education & #healthcare programs http://cap.af/gARZd1 #p2 #budget"*	*"#Socialimpactbonds predicted to be worth $billions within 20yrs http://t.co/rrl7zhl I agree!"*
Neutral	*"We're still looking for video questions for @slashklc & @daniel_stid for today's #socialimpactbonds show. Record here: http://ow.ly/cPXwU"*	*"After seven years 'in the trenches' as part of Australia's foray into social impact bonds, Elyse Sainty outlines her 'thoughts and prayers' for the future of social impact bonds. #impinv #socialimpactbonds"*	*"Have questions about #socialimpactbonds? Use #rcsocialimpact to ask #hks prof Liebman"*	*"What r #socialimpactbonds (AKA #SIB) @DLeonhardt explains in february 2011 @nytimes article http://t.co/NUKde4bF"*	*"New York Times reviews emergence of #socialimpactbond http://nyti.ms/Rn8w7H #socent #socinn"*
Negative	*"SIBs uphold poverty and exclusion (systemic violence). SIBs do not abolish harmful social/economic constructions. Cut this crap, fight the SIBs! https://doorbraak.eu/social-impact-bonds-disciplinary-system-against-people-at-the-bottom-of-society/ #socialimpactbonds #philanthropy #capitalism #poverty #violence"*	*"#MinistryofJustice sold £5m of #SocialImpactBonds to philanthropists. #Capitalist guilt money? http://t.co/az4H3SI #tiltingatwindmills #in"*	*"#SocialImpactBonds are not always the answer for social problems. Intervention without real knowledge about the root cause of a problem in a community can cause more negative consequences than benefits. #socialimpact https://t.co/GGO3zDqJRv"*	*"Polly Toynbee's ferocious attack on #socialimpactbond: It's an illusion to expect to make money tackling social problems http://t.co/57N7Joa"*	*"Making profit out of despair and deprivation makes me want to puke. #socialimpactbonds for investors. Tackle the cause #incomeinequality."*

While analyzing the sentiments contained in SIB-focused Tweets in isolation provides some interesting insights, it is essential to also consider how the sentiments have differed from those expressed in other forms of data, including academic publications and other social

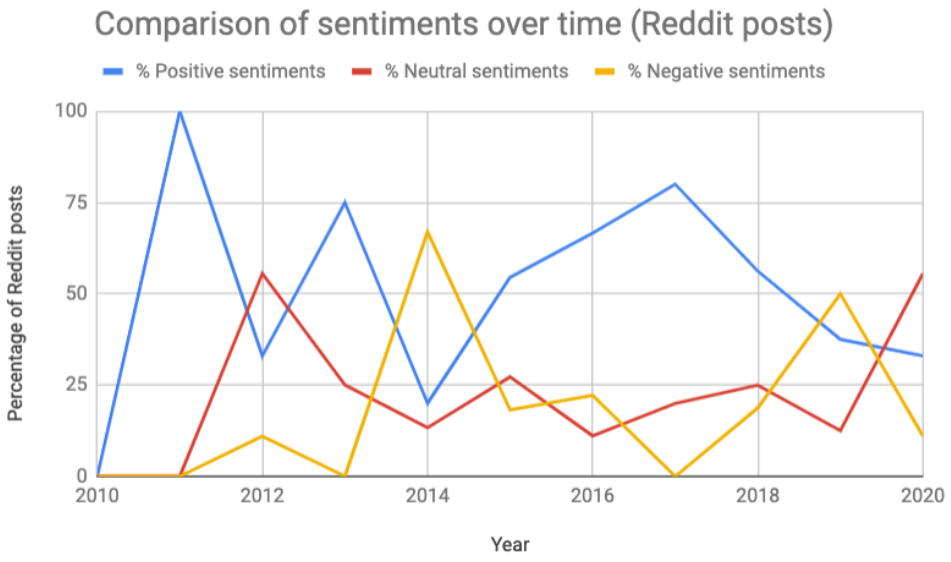
media data (e.g. Reddit posts). Figure 7.20 thus visualizes how sentiments differed in academic publications between 2010 - 2020 (including papers published in peer-reviewed journals that are focused on SIBs). The academic papers were coded for sentiment by reading the discussion and conclusion section and labelling them according to positive, neutral or negative sentiment. Interestingly, the positive sentiments were growing in academic articles, while they were dipping in tweets; and they have declined in recent times (in the academic literature), while positive sentiments in tweets have increased. Likewise, the trends for negative sentiments are quite different, with negative sentiments appearing in the very early publications, then disappearing; and steadily increasing since 2014. Methodologically, this supports the earlier argument about the citations data lagging behind social media data and the importance of extending the analysis beyond citation tracking. In terms of my research question, these patterns provide further support for the institutional view, as one would expect to see a closer match between results, sentiments and uptake if SIBs were adopted for instrumental reasons.

Figure 7.20: Sentiments Expressed in Academic Literature Over Time



Differing patterns were also observed when comparing Twitter sentiments to those expressed in other forms of social media data, such as Reddit data (see Figure 7.21). Reddit is a centralized community platform categorized around different topics; there are more than 10,000 anonymous online communities known as "subreddits" (Sowles et al., 2018; van der Nagel & Frith, 2015). Unlike other social media sites, Reddit promotes the use of "throwaway accounts" (i.e. a temporary account used for a specific purpose) and pseudonyms to facilitate anonymous discourse (van der Nagel & Frith, 2015). Posts from "throwaway accounts" are associated with increased negativity, thus making Reddit essential to consider when identifying more critical viewpoints about SIBs (van der Nagel & Frith, 2015). Several authors have noted that Reddit contains expansive, high-quality content (Medvedev et al., 2019; Zamani et al., 2019). It has thus been used to investigate perceptions of specific phenomena (e.g. Brett et al., 2019; De Choudhury & De, 2014).

Figure 7.21: Comparison to Sentiments Found in Reddit Posts From 2010 - 2020



As highlighted in Figure 7.21, positive sentiments have declined in Reddit posts since 2017 (while increasing in tweets). Prior to that, they increased from 2014 - 2017 (while decreasing in tweets); and varied between 2010 to 2014. Negative sentiments were

exceptionally high in 2014 and rose markedly again in 2019. Apart from those years, they have featured much less than positive sentiments, similar to the Twitter dataset. Table 7.5 consolidates the data on changing sentiments, comparing the positive sentiments contained in Tweets, academic papers and Reddit posts by year. As expected, the volume of Reddit posts and academic papers was far lower than that of Tweets; given the small sample size of these data sources, the results should be interpreted cautiously.

Table 7.5: Comparison of Positive Sentiments (by year) - Tweets, Reddit Posts, Academic Papers

Year	% Tweets coded positive (sentimentR) (#)	% Reddit posts coded positive (#)	% academic papers coded positive (#)
2010	53% (n = 10)	N/A (n = 0)	N/A (n = 0)
2011	67% (n = 307)	100% (n = 3)	33% (n = 1)
2012	70% (n = 797)	33% (n = 3)	N/A (n = 0)
2013	61% (n = 1267)	75% (n = 3)	44% (n = 4)
2014	45% (n = 836)	20% (n = 2)	57% (n = 8)
2015	43% (n = 915)	54.50% (n = 6)	56% (n = 5)
2016	48% (n = 919)	66.66% (n = 6)	52% (n = 11)
2017	48% (n = 524)	80% (n = 8)	43% (n = 10)
2018	64% (n = 1077)	56.25% (n = 9)	29% (n = 12)
2019	63% (n = 648)	37.50% (n = 3)	27% (n = 11)
2020	63% (n = 269)	33% (n = 3)	23% (n = 12)

R2: Which stakeholders have played a central role in communicating and promoting the SIB idea in online forums such as Twitter?

To identify the most central nodes (i.e. people in the network) in terms of the different dimensions of centrality (i.e. how important actors are in a network's structure), the following measures were calculated: 1) in-degree score; 2) out-degree score; 3) eigenvector; and 4) betweenness centrality measure (Borgatti et al., 2009; Borgatti et al., 2013). Each measure was described in more detail in the methods section of this chapter (e.g. Table 7.2). In

presenting the top centrality scores (Tables 7.7 – 7.10), I also note whether the relevant account has direct SIBs involvement. In this section, I define 'direct involvement' as including consultants and those receiving the consultancy and implementing the SIB. The GO Lab's database provides a directory of involved stakeholders (Go Lab, 2023d). While this assisted with understanding whether an individual or organization has a vested interest in the success of a SIB, it must be noted that it can be difficult to distinguish between 'involved' or 'non-involved' stakeholders in the context of SIBs. For instance, while the GO Lab is not listed on the INDIGO database, they are evaluating the Life Chances Fund SIBs (Fitzgerald et al., 2022) and are 'involved' in an advisory capacity across the sector. Understanding the extent to which basic network measures are robust is vital for assessing the validity of network research (Borgatti et al., 2006).

The selected centrality measures provide a strong starting point for analysis as they are among the most frequently used network measures. Thus, significant attention has been directed towards understanding their robustness. Borgatti et al., (2006) systematically explored the robustness of centrality measures in the context of different amounts and types of measurement error. They found that the measures tested (degree, betweenness, closeness and eigenvector) were reasonably robust under small amounts of error (10% and under). Thus, even if the data collection method missed 5% of the ties (connections between people) in a network, the correlation between true and observed centrality will be in the .90s; and the probability of correctly identifying the most central node is around 90%. While there are various study limitations (e.g. Borgatti et al., 2006 analyzed random error on random networks), this suggests that well-known centrality measures can provide quite robust results even with small amounts of error in the data. In the SIBs context, this is likely to be sufficient, while in other contexts (e.g. an epidemic where failing to vaccinate the most

central actors could cost lives), this may not be a reliable enough proxy (Borgatti et al., 2006; Borgatti, 2003).

Table 7.6 outlines the top ten stakeholders in the network according to in-degree score. Users with a high in-degree score gain attention from their tweets among the other actors in the network; the score captures the number of links others have initiated with the user (e.g. by mentioning them) (Himelboim et al., 2017; Priyanta., & Trisna, 2019). Overall, many tweeters who are 'conversational hubs' (thus having a high in-degree score) have a high level of direct involvement in SIBs, with stakeholders from the "fashion setting network" well represented. This provides further evidence of SIBs diffusing for reasons other than efficiency gains, as those who are highly connected and whose information cascades throughout the network are largely stakeholders who can benefit financially and/or reputationally from the success of SIBs. Below, I provide some specific examples that further contextualize these broad observations.

The account that received the highest in-degree score, thus indicating the most mentions, retweets or replies, is Social Finance UK, a not-for-profit consultancy organization. Social Finance UK is directly involved in 25 SIBs (as an intermediary) and thus has a direct stake in the success of the SIB idea. More specifically, several prominent Co-Founders of Social Finance (e.g. Toby Eccles and Ronald Cohen) have been heralded as the individuals who developed the SIB idea. After Social Finance UK, the account with the second highest in-degree score is an individual (Ruben Koekoek) who also has strong ties to Social Finance (and thus the SIB concept) as the Co-founder and Managing Director of Social Finance Netherlands. Social Finance US is in the fourth position, suggesting that the US arm of the organization has played a particularly influential role in cascading information about SIBs

throughout the network. Out of the top 10 accounts, seven have direct involvement with SIBs, demonstrating that the concept has been actively promoted by stakeholders who directly benefit from its success. The other accounts with a high in-degree score include: 1) Centre for SIBs (based in the Department for Culture, Media and Sport; co-commissioner for 39 SIBs); 2) The Rockefeller Foundation (Investor in NYC Increasing Employment and Improving Public Safety SIB); 3) Michael Eddy (Founding Partner of Instiglio; intermediary in 5 SIBs); and the OPM Network (evaluator for the Essex SIB). In addition, the GO Lab features in the top 10 accounts; they were created as a partnership between the Blavatnik School of Government and the UK Government, and have conducted extensive research on the use of SIBs, including evaluations (FitzGerald et al., 2021). The distribution suggests that intermediaries/consultants have been particularly influential in spreading information about SIBs. Service providers did not feature in the top 10 users according to in-degree, and commissioners were also under-represented apart from the Centre for SIBs. Investors and evaluators/researchers played a minor role.

Table 7.6: Top 10 Users According to In-Degree Score

Twitter username	Name	Type of account	Stakeholder Category	Direct SIBs involvement?	# of SIB projects directly involved with	In degree score
socfinuk	Social Finance UK	Organization	Intermediary	Yes	25	207
rubenkoekoek	Ruben Koekoek	Individual	Co-founder and Managing Director of Intermediary	Yes	6	144
socimpactbonds	Centre for SIBs	Organization	Government Department	Yes	39	143
SocialFinanceUS	Social Finance	Organization	Intermediary	Yes	8	133
RockefellerFdn	The Rockefeller Foundation	Organization	Investor	Yes	1	129
HUMANFdn	Human	Organization	Think Tank	No	0	117

	Foundation					
FinanceforGood	Finance for Good	Organization	Intermediary	No	0	103
MichaelEddy	Michael Eddy	Individual	Consultant/Director at Investment Firm	Yes	8	97
OPMnetwork	OPM Network	Organization	Evaluator	Yes	1	94
ukgolab	Government Outcomes Lab	Organization	Research institution	No	0[6]	88

Table 7.7 displays the top 10 users with the highest out-degree score. Users with a high out-degree score initiate a high level of engagement with the community, for example, by tweeting a lot about a particular topic and mentioning and/or replying to other users (Himelboim et al., 2017; Jastania et al., 2020; Schuster & Kolleck, 2020). Slightly fewer organizations (five compared to seven for the in-degree score) have direct SIBs involvement. Regardless, this measure also provided evidence of institutional drivers, given that stakeholders in the "fashion setting network" (e.g. intermediaries) were once again well represented, while stakeholders with an interest in evidence and/or who are involved in generating evidence (e.g. evaluators, academics) were severely underrepresented.

The most active user was Social Finance UK (which also had the highest in-degree score), an intermediary involved with 25 SIBs. Other particularly active accounts belonged to intermediaries (Finance for Good, NFF Social Impact, Social Finance US, Social Impact Finance and Ruben Koekoek – Co-Founder of Social Finance Netherlands). To a lesser extent, researchers/think tanks featured, for example, the GO Lab and Giovanna Melandri (President of a think tank). A similar number of accounts (to the number observed for the top

[6] The INDIGO Organisation Directory does not list the GO Lab as being directly involved (i.e. involved in 0 projects), although it is known that they have evaluated several SIB projects. The total number of projects is unclear, however, due to the number being omitted in the INDIGO database.

in-degree scores) were organizational rather than individual accounts. This suggests that organizations with a broad reach have primarily been the most active in initiating discussion in the network. Multiple accounts connected to the same organization (Social Finance) featured among the top out-degree scores (as with the top in-degree scores).

Table 7.7: Top 10 Users According to Out-Degree Score

Twitter username	Name	Type of account	Stakeholder Category	Direct SIBs involvement?	# of SIB projects directly involved with	Out degree score
socfinuk	Social Finance UK	Organization	Intermediary	Yes	25	55
FinanceforGood	Finance for Good	Organization	Intermediary	No	0	55
ukgolab	Government Outcomes Lab	Organization	Research institution	No	0	53
GioMelandri	Giovanna Melandri	Individual	President of think tank	No	0	48
NFFSocialImpact	NFF Social Impact	Organization	Investor/intermediary	Yes	5	41
SocialFinanceUS	Social Finance	Organization	Intermediary	Yes	8	39
InvisHeartFilm	The Invisible Heart	Orgsanisation	Documentary	No	0	37
MattGripp	Matt Gripp	Individual	Attorney	No	0	34
socialimpactfin	Social Impact Finance	Organization	Intermediary	Yes	1	31
rubenkoekoek	Ruben Koekoek	Individual	Co-founder and Managing Director of Intermediary	Yes	6	29

Table 7.8 presents the top ten accounts according to betweenness centrality. Users with a high betweenness score are highly connected, performing a vital mediation role (Zhang & Luo, 2017). They typically have the shortest path between the vertices of the graph (i.e. as more people depend on a user to make connections, the users' betweenness centrality score becomes higher), and they can thus control information flow around the network

(Hevey, 2018). This score is particularly relevant for my research question as actors with a high betweenness score are integral in forwarding ideas and information and can shape the concept in favour of their interests. In each of the measures presented so far (including betweenness centrality), service providers have not featured in the top accounts. Evaluators have featured, although they are very underrepresented; a government department has only featured in the top 10 scores for in-degree and betweenness centrality. The near absence of stakeholders involved with generating evidence again suggests institutional drivers to reform adoption have been more prominent than instrumental drivers, with more of a focus on rhetoric rather than results.

The most central account in the SIBs network was Social Finance UK; as previously mentioned, they are a consulting firm directly involved with 25 SIBs as an intermediary (they also had the highest in-degree and out-degree scores). As with Social Finance UK, most of the accounts in the top 10 for betweenness centrality have already been mentioned (due to ranking highly for in-degree and/or out-degree scores). For instance, Finance for Good, Social Finance US, NFF Social Impact, Ruben Koekoek, Centre for SIBs, The Invisible Heart, Human Foundation and the GO Lab. The only account yet to be featured is Impact Investing at MaRS, an account connected to an intermediary organization involved in 2 SIBs. As with the other categories, most accounts identified as having a high betweenness centrality score are owned by organizations rather than individuals. Consultants/ intermediaries were once again particularly influential; this is notable for this category, in particular, as the score identifies ‘brokers’ between different parts of the network.

Table 7.8: Top 10 Users According to Betweenness Score

Twitter username	Name	Type of account	Stakeholder Category	Direct SIBs involvement?	# of SIB projects directly involved with	Betweenness score
socfinuk	Social Finance UK	Organization	Intermediary	Yes	25	177,156.0
FinanceforGood	Finance for Good	Organization	Intermediary	No	0	147,196.6
SocialFinanceUS	Social Finance	Organization	Intermediary	Yes	8	79,955.4
NFFSocialImpact	NFF Social Impact	Organization	Investor/intermediary	Yes	5	73,136.0
rubenkoekoek	Ruben Koekoek	Individual	Co-founder and Managing Director of Intermediary	Yes	6	65,988.6
socimpactbonds	Centre for SIBs	Organization	Government Department	Yes	39	63,579.3
socialfinance	Impact Investing at MaRS	Organization	Intermediary, provider, investor	Yes	2	58,332.1
InvisHeartFilm	The Invisible Heart	Orgsanisation	Documentary	No	0	55,623.3
HUMANFdn	Human Foundation	Organization	Think Tank	No	0	45,304.2
ukgolab	Government Outcomes Lab	Organization	Research institution	No	0	44,605.4

The distribution of eigenvector centrality (Table 7.9) also provides crucial insights into the actors' roles in the overall diffusion network. This score is assigned based on scores that neighbours have received, thus recognizing that connections to more influential vertices are more critical than those with less influence (Golbeck, 2013). Ultimately, *"a position is central, with respect to eigenvector centrality, if it is connected to other positions which have many connections"* (Bienenstock & Bonacich, 2021, p. 232). Table 7.9 demonstrates that the accounts that are particularly connected (thus having a high eigenvector score) are dominated by stakeholders with direct SIBs involvement (9 out of the top 10 accounts). This again suggests that SIB stakeholders with a vested interest have actively shaped and spread the idea

online; notably, SIB adopters are highly connected to influential ‘neighbours’ in the network, enabling information to spread easily and quickly from and between SIB adopters.

As outlined in Table 7.9, critical actors in the field held the most central positions regarding eigenvector score, particularly intermediaries such as Instiglio. Several previously mentioned actors also have a high eigenvector score. Regardless, it is worth noting that the distribution looked slightly different compared to other centrality measures and that some particularly influential stakeholders (e.g. Social Finance) were not featured in the top 10 for this category. Organizations rather than individuals once again owned the majority of the accounts (seven out of 10). Intermediaries/consultants were again the most common stakeholder type (5 out of the top 10), although there was more of a spread of other stakeholders (including a Government Department, an evaluator, investors and an innovation laboratory). The featured organizations included Instiglio (intermediary for 5 SIBs); Finance for Good (no SIBs involvement); Centre for SIBs (co-commissioner for 39 SIBs); The Rockefeller Foundation (investor in 1 SIB); NFF Social Impact (investor in 5 SIBs); IDB - Inter-American Development Bank (investor in 3 SIBs): and Traverse (formerly OPM, evaluator in 1 SIB). The only individuals included in this list were Michael Eddy (Founding Partner of Instiglio, intermediary in 5 SIBs); Emma Tomkinson (who has worked on SIB policy for the NSW Government and UK Cabinet Office); and Johannes Lohmann (Head of Work & Finance at Behavioural Insights Team; previously CFO of Instiglio – intermediary for 5 SIBs; and has past involvement with ICF, the World Bank's investment firm – commissioner for 2 SIBs). The existence of individual accounts that can be linked to organizations (e.g. Johannes Lohmann previously worked for Instiglio) further reinforces the central position of many organizations.

Table 7.9: Top 10 Users According to Eigenvector Score

Twitter username	Name	Type of account	Stakeholder Category	Direct SIBs involvement ?	# of SIB projects directly involved with	Closeness score
InstiglioInc	Instiglio	Organization	Intermediary	Yes	5	1
MichaelEddy	Michael Eddy	Individual	Consultant/Director at Investment Firm	Yes	8	0.6715247266
FinanceforGood	Finance for Good	Organization	Intermediary	No	0	0.3917899994
socimpactbonds	Centre for SIBs	Organization	Government Department	Yes	39	0.2211960112
emma_tomkinson	Emma Tomkinson	Individual	Consultant	Yes	Unclear	0.2168240267
RockefellerFdn	The Rockefeller Foundation	Organization	Investor	Yes	1	0.2079520952
NFFSocialImpact	NFF Social Impact	Organization	Investor/intermediary	Yes	5	0.2030884006
hanneslohmann	Johannes Lohmann	Individual	Consultant - Behavioural Insights Team	Yes	6	0.1794339751
IDB_Lab	IDB Lab	Organization	Innovation laboratory	Yes	3	0.1499641631
traversepeople	Traverse	Organization	Evaluator	Yes	1	0.132199565

7.6 Discussion

This chapter considered how sentiments regarding SIBs have changed over time on Twitter and the extent to which the sentiments expressed in tweets have reflected emerging (inconclusive) research findings. The analysis found that the dataset overwhelmingly featured positive and neutral tweets, with negative tweets constituting a much smaller share. This finding is consistent with existing Twitter studies which have highlighted that while tweets do not always provide a source of support or positive messaging (Wynn et al., 2017), negative tweets are typically less common (Welch & Mihalcea, 2017). For instance, Misuraca et al., (2020) applied the five main R packages used for sentiment analysis to the Student Course Evaluation Comments dataset built by Welch & Mihalcea (2017). They found that

tweets with negative sentiments (ranging between 15.72% - 33.57% depending on the package) were less common than tweets with positive sentiments (ranging between 57.07% - 75.62%). Negative tweets were more severely under-represented in the SIBs dataset than in the Misuraca et al., (2020) study, dipping as low as 4.1% for the meanR package in 2017. A higher proportion of neutral sentiments led to the proportion of positive SIB tweets being slightly lower in this study compared to the Misuraca et al., (2020) study. The existing Twitter literature provides insights into why negative tweets are likely to be over-represented in financing contexts, citing difficulties with classifying words commonly used in a financial context. For instance, using a large sample of corporate reports from 1994 – 2008, Loughran and McDonald (2011, p. 35) found that *"almost three-fourths of the words identified as negative by the widely used Harvard Dictionary are words typically not considered negative in financial contexts".* Nonetheless, in the SIBs context, negative tweets remained severely under-represented, suggesting that the tweets potentially lacked technical financial terms that could have easily been misclassified as negative, and instead focused on the broader benefits.

It is also integral to consider broader trends regarding internet usage when interpreting the patterns identified in this chapter. Increased access to technology, the modernization of various countries, and increased smartphone utilization have enabled a greater proportion of people to use the internet more easily and frequently (Statista, 2021b). As a result, the number of internet users has continued to increase; for instance, in 2021, the number of global internet users was 4.9 billion, compared to 4.6 billion in 2020 (Statista, 2021b). In addition, the percentage of the global population accessing the internet has increased. As of April 2022, there were 5 billion internet users worldwide, equating to 63 per cent of the global population (4.65 billion internet users were social media users) (Statista, 2022). The

online decline in SIBs interest across most platforms, is thus particularly stark when considering the broader internet usage trends.

My first research question asked how sentiments regarding SIBs have changed over time and the extent to which the sentiments contained in tweets express or reflect emerging (inconclusive) research findings. The findings related to this question are consistent with the existing literature, which highlights that policy advocates often oversell policy reforms, by emphasizing benefits while down-playing or ignoring critiques (Brunsson & Olsen, 1993; King & Crewe, 2014). Regardless of which pre-trained model was applied, the tweets were overwhelmingly positive throughout the entire time period. This was also consistent with the manual coding I conducted on the entire dataset (as depicted in Figures 7.12 – 7.15). Positive sentiments increased between 2010-2012 before dipping between 2012-2016 and steadily increasing again from 2017-2020. The citation data lagged behind social media data, with positive sentiments growing in academic articles while they were dipping in tweets and vice versa. Both Twitter and Reddit posts were far more positive than academic publications, even though Reddit posts are often associated with increased negativity (van der Nagel & Frith, 2015). This demonstrates that the proportion of negative tweets (and likely posts on other social media platforms) has not grown as emerging (inconclusive and sometimes negative) research findings have accumulated and that there is strong evidence of adopters enthusiastically promoting SIBs despite limited evidence of effectiveness. Overall, the sentiment patterns identified in this chapter provide evidence of the existence of pro-innovation biases, along with institutional motivations for adoption (Abrahamson, 1991).

The second research question considered which stakeholders have played a central role in communicating and promoting the SIB reform, in online forums such as Twitter. The

analysis found that many of the most central actors in the network have direct SIBs involvement, often in a role that enables them to profit off the success of SIBs (e.g. as an intermediary or investor). This was consistent across various centrality measures. These findings are aligned to the theoretical work by Blumer (1969, p. 595), who noted that fashion setters play an active role in *"selecting a few administrative models from many and in developing organizations' awareness and tastes for these models in order to render them fashionable and to prompt their diffusion".* Additionally, the findings support the work of various organizational theorists, who have claimed that stakeholders such as consultants, business schools and business mass media tend to populate fashion-setting networks (Abrahamson, 1986; DiMaggio & Powell, 1983; Kimberly, 1981; Mintzberg, 1979). Specifically, Abrahamson (1991, p. 603) noted that *"when fashion-setting networks back the diffusion of technically inefficient innovations, they will tend to promote their faddish diffusion".* While the Twitter data seemed to support the notion that "fashion setting" organizations have played a vital role in the diffusion of the idea, it is essential to recognize that a wide array of stakeholders can benefit from SIBs to differing degrees in different ways. Thus, it can be more challenging to categorize stakeholders as 'involved' or 'not involved' or 'fashion-setting' versus 'non-fashion-setting' than in other settings. For example, it is not only intermediaries and investors that profit from SIBs; even when only considering financial benefits, organizations such as evaluators and research organizations (e.g. GO Lab) stand to gain from continued SIBs growth, as do service providers interested in future contracts.

The analyses are limited in various ways, many of which will be overcome in the next chapter. Firstly, the sentiment scores calculated by the pre-built packages assigned a score based on the content of the entire tweet; this may or may not have aligned with specific attitudes about SIBs. For instance, a tweet would have likely been automatically coded as

positive if it enthusiastically mentioned the launch of a paper, even if the referenced paper contained negative sentiments. Thus, the automated approach is more scalable but less accurate than manual coding. Secondly, some SIB-specific tweets may have not have been retrieved due to the selection of hashtags or due to tweeters not assigning a relevant hashtag when tweeting about the topic. Thirdly, there are Twitter-specific biases and limitations, such as the fact that: 1) Twitter user demographics are not overly representative; 2) as some people tweet more often than others, a Twitter dataset is likely to over-represent the opinions of a small subset of users; 3) tweets posted by bots can be hard to detect and can skew the data and individuals may also try to manipulate Twitter algorithms for their own purposes; 4) tweets that rely on links and images can be difficult to accurately analyze using sentiment analysis libraries; and 5) detecting sarcasm and humour (common in tweets) can be difficult when using natural language processing approaches (Cihon & Yasseri, 2016). Additionally, the impact of COVID-19 on the patterns identified in the data remain unclear as data from 2021 onwards were not incorporated into the dataset for pragmatic reasons.

While this chapter analyzed the sentiments contained in the entire dataset using pre-built sentiment libraries (with a comparison to sentiments coded manually), it did not involve rich qualitative analysis. Thus, it is unclear which persuasive techniques have been used, by which stakeholders, to convince decision-makers to adopt the SIB innovation (Hood & Jackson, 1991). Furthermore, in analyzing which actors have played a central role in the SIB network, the extent to which each actor has a COI was not considered systematically with a scale (Eisner et al., 2015). This is a potentially significant omission to be overcome in the next chapter, given that a wide array of stakeholders are involved in a SIB contract, and the extent of their vested interests varies considerably.

7.7 Conclusion

This chapter has contributed to understanding how and why the SIB innovation has diffused throughout various public sectors, by analyzing social media data. While some studies in this field have analyzed the perspectives conveyed in the academic and grey literature, the social media data have, so far, not been incorporated into academic studies of SIBs. Additionally, no attempts have been made to identify and understand the networks formed on platforms such as Twitter, and the ways in which particular SIB stakeholders have shaped the online discourse. This is a significant void to fill, given that SIBs have been socially constructed over time, and many of the key SIB stakeholders who have actively worked on one or more SIBs, have regularly tweeted about the idea, thus helping to spread the idea further.

This chapter has demonstrated that positive tweets overwhelmingly dominate Twitter content, while negative tweets are under-represented. As hypothesized, as inconclusive (and sometimes negative) findings have accumulated in the academic literature, there has not been a similar growth in stakeholders tweeting negatively about the innovation. Thus, minimal attention has been directed to the limitations, critiques and unintended consequences associated with SIBs. This is concerning as many social interventions have caused harm despite the best intentions. Without critical discourse, it is possible that pro-innovation biases are being perpetuated in online forums such as Twitter (Abrahamson, 1991; Bonell et al., 2015; McCord, 2003). In considering which stakeholders have played a central role in propelling the diffusion of SIBs online, it was apparent that fashion-setting organizations such as consulting firms have very actively promoted the idea on Twitter. In addition, the most influential users, across various indicators of network centrality, were highly involved in SIB projects, often in an intermediary capacity. Chapter 8 will build upon this analysis by

considering which discourses have been heavily promoted; and which persuasive techniques have been used by SIB stakeholders. When combined, the findings of this chapter and Chapter 8 will provide essential insights into how and why the idea has spread, despite limited evidence of its effectiveness.

Chapter 8: Framework Analysis of Social Media Data

8.1 Introduction

The previous chapter identified that reform "chatter" on Twitter heavily featured positive and neutral sentiments regarding SIBs, compared to negative sentiments, and this was apparent across all periods. It noted that positive sentiments increased steadily between 2017-2020 (after dipping between 2012-2016); however, negative sentiments increased in the academic literature during this time. Thus, as initially hypothesized, as inconclusive (and sometimes negative) findings have accumulated in the academic literature (Chapter 5, Figures 5.4-5.5), there has not been a similar growth in stakeholders tweeting negatively about the innovation. Further, the previous chapter identified evidence of organizations and individuals within the "fashion setting community" (particularly consultants) promoting SIBs in the Twitter discourse. These findings provide evidence of institutional drivers, although the signals are mixed and require further investigation. Just as it was necessary to explore sentiments rather than only the volume of tweets, it is essential to further unpack the content of the tweets, using rich qualitative methods. Therefore, this chapter primarily uses framework analysis to better understand the discourses and themes in the Twitter dataset.

The aim is to identify how persuasive techniques have been used by SIB stakeholders to frame the idea as a 'legitimate' reform to pursue over time (Hood and Jackson 1991); and which discourses have been particularly prominent. While I used the terms 'adopter' and 'non-adopter' in the previous chapter, I will refer to 'directly involved', 'broadly involved' and 'uninvolved' stakeholders in this chapter. Differing terminology is required as some of the stakeholder categories that are included in the dataset (e.g. media, politicians etc) could not have been contractually involved in a SIB. However, they have had some level of involvement in spreading the idea. Recognizing that there are differing levels of stakeholder

involvement and vested interests, I also categorize Twitter accounts by the level of conflict of interest (COI). I use the term COI to refer to the extent to which a stakeholder would receive financial and/or reputational benefits if the SIB idea succeeds, thus leading to an increased level of bias (Eisner & Humphreys, 2012).

In combination, the analyses included in this chapter build on the previous chapters and provide further (and much deeper) insights into why the idea has diffused throughout many public sectors since 2010. While the prior chapter considered broad trends (e.g. volume of tweets over time) and assessed the sentiments and network centrality measures for the entire dataset, the (largely quantitative) methods used did not allow for in-depth consideration of the content of the tweets. In this chapter, the analyses are presented in two components, one involving a random sample and another involving a purposive sample. The results stemming from the random sample demonstrated that: 1) the persuasive techniques identified by Hood and Jackson (1991) are highly evident in the tweets, suggesting it is an appropriate framework for the analysis of social media data; 2) positive sentiments appear to be more prevalent for involved stakeholders, however, it does not appear that these positive opinions are based on evidence; and 3) COI level appears to influence sentiment, with those who have a medium or high level of COI expressing a higher percentage of positive sentiments about the reform idea. Regarding the component that uses a purposive sample, the findings demonstrate that: 1) the discourse categories identified in a prior SIBs study (Ormiston et al., 2020) were well-represented in the Twitter dataset, although some additional categories (focused on harm, excessive hype etc) were identified; 2) robust academic studies and impact evaluations were rarely mentioned in the top retweets; 3) and the ‘events’ mentioned suggest that we often set things up with great enthusiasm and do not adequately consider the results.

8.2 Hypotheses Development

As with the previous chapter, the data reported in this chapter address two overarching research questions, each focused on gathering further evidence on possible institutional drivers.

R1: What persuasive techniques have been used by involved stakeholders to frame the SIB funding model as a 'legitimate' reform to pursue over time?

Related hypotheses: *The 'keys to persuasion' outlined in the public administration literature will feature highly in the tweets, particularly those posted by stakeholders who have some level of involvement with the reform.*

The superseding of one idea over another in public administration often *"occurs through a rhetorical process, not by the marshalling of incontrovertible evidence from exhaustive examination of data"* (Hood & Jackson 1991, p. xi). It is through this rhetorical, social process that some ideas 'win' over potential rivals, with competing ideas often being treated as *"heresies or outdated ideas"* (Hood & Jackson 1991, p. 17). It is expected that this will be reflected in the Twitter data, with involved stakeholders resorting to well-accepted strategies of persuasion, that have been used widely, in public administration, to promote ideas. If evidence of this is identified, it could suggest the existence of institutional drivers, as the debate (and potentially adoption decisions) would be primarily centred around rhetoric, norms, and beliefs rather than a quest for efficiency (Greenwood et al., 2017).

R2: What discourses can be identified in the highly shared tweets; and to what extent have particular discourses dominated/others been drowned out?

Related hypotheses: *a high percentage of the highly retweeted tweets will focus on possible future benefits and assume that adoption is beneficial, while downplaying or ignoring any disadvantages.*

The public administration literature emphasizes that policy advocates often over-sell the benefits of reforms and fail to robustly measure them (Baier et al., 1986; Brunsson & Olsen., 1993; Hood & Dixon, 2015; King & Crewe., 2014). Thus, it is anticipated that the most highly shared tweets will be overly optimistic about SIBs, with minimal discussion of SIB evaluations or academic studies. Related to this, as SIBs have been socially constructed by different stakeholders with different needs and interests, it is expected that a variety of discourses will feature in the tweets (Fraser et al., 2020, 196; Greenwood et al., 2017; Ormiston et al., 2020). The discourse categories identified in a past SIB discourse analysis study (Ormiston et al., 2020) are expected to feature in the retweets; although it may be possible to identify extra categories, given that the rhetoric featured in Tweets is likely to differ considerably to that found in media releases. When analyzing the discourses, it will be essential to consider how decoupling (i.e. the gap between policy and practice) might influence the extent to which the discourses found in tweets are congruent with real-world practice (Bromley & Powell, 2012).

8.3 Overview of Methods

This chapter primarily uses rich qualitative methods to build on the largely quantitative methods and findings of the previous chapter. There were two main components to the analysis presented in this chapter. Firstly, a random (10%) sample of the cases (original tweets only) was constructed using R. The random sample consisted of 742 tweets, including tweets posted by ‘involved’ and ‘uninvolved’ stakeholders, from different time points and different jurisdictions. To qualitatively analyze the random sample, I used framework

analysis. This involved applying Hood and Jackson's (1991) framework (regarding strategies of persuasion) to the random sample to identify the extent to which persuasive techniques are evident in the data. I also created a scale that measures the extent to which each tweeter has a COI, and I analyzed the impact of COI on sentiments. Secondly, I created a purposive sample of the most highly retweeted tweets, obtained by resampling the whole dataset and sorting the data (in R) based on descending order of retweet counts (283 qualifying tweets, accounting for 2,211 retweets). For this component, I also used framework analysis, which involved applying an existing framework developed in a recent discourse analysis of SIB-related press releases (Ormiston et al., 2020). I identified several additional (mainly cautionary) discourse categories, which were primarily focused on the excessive hype and potential harm surrounding SIBs. The methods for each of these two components are described below.

Overarching approach - framework analysis

Framework analysis was used to better understand what Twitter data reveals about how and why SIBs have rapidly spread throughout the public sector. It is an appropriate qualitative approach as there are pre-existing frameworks in the public administration literature that are relevant and worthy of empirical examination. As recommended by Ritchie and Spencer (1994), I followed five stages of framework analysis, including: 1) data familiarization (reviewing a subset of the tweets and making initial notes about key ideas); 2) framework identification (constructing a framework for the analysis and resulting interpretation, drawing on a priori assumptions and concepts); 3) indexing (systematically applying the framework to all of the study data); 4) charting (ordering the indexed data; in a matrix form, which uses ordered rows and columns populated by the framework components and units of analysis); and 5) mapping and interpretation (combining the key learnings from

the earlier steps) (Ritchie & Spencer, 1994). Below, I describe the framework that was adopted for each part of the analysis.

Strategies of persuasion framework – random (10%) sample

To understand why SIBs have, to use Hood and Jackson's terminology, 'fitted the lock' (i.e. been accepted) fore more than 10 years, as well as why usage has appeared to decline recently, I operationalized Hood and Jackson's (1991, p. 184) 'acceptance factor' items. These include "*symmetry, metaphorical force, ambiguity, private benefit successfully presented as public good, selective evidence rather than the production of irrefragable proofs from systematic analysis of hard data, and suspension of disbelief".* While some of the elements in Hood and Jackson's (1991) framework (e.g. ambiguity and selectivity of evidence) were difficult to operationalize and thus excluded, I included most of their components in the framework. I identified illustrative examples for each of the strategies of persuasion. I also coded some additional elements (listed under "background coding" in Table 8.1) to allow for more nuanced analysis.

Table 8.1: Coding Framework – Persuasive Techniques

	Item	***Description***
Background coding	Stakeholder type	Broad stakeholder category (e.g. media, research, finance etc)
	SIBs involvement	Whether directly involved with a SIB contract (used INDIGO database as a list of involved organizations)
	Alternatives to SIBs discussed	Whether an alternative funding instrument was mentioned
	SIB evaluation mentioned	Whether a specific impact evaluation linked to a SIB was mentioned
	Sentiment contained in tweet	Labelled as positive, neutral, or negative – depending on thoughts/feelings expressed towards SIBs
Public admin persuasive techniques	Symmetry	Whether the tweet referred to a problem (e.g. lack of funds for early intervention) and mentioned SIBs as the solution
	Non-literal – metaphors, imagery etc	Whether the tweet included metaphor, humour or imagery
	Emphasizing public good	Whether the tweet justifies private benefit in terms of public good
	Suspension of disbelief – framing idea as new	Whether the tweet referred to the SIB idea being new (e.g. language such as 'groundbreaking', 'innovative', a 'first' etc)
	Suspension of disbelief – character assassination	Whether tweet questions the character of those who oppose SIBs
	Suspension of disbelief – time pressure	Whether the tweet appeals to time urgency
	Suspension of disbelief – linking to	Whether SIB directly links national economic performance

	national economic performance	with the quality of public management
Twitter-specific persuasive techniques	Use of multiple hashtags	Whether the tweet contained multiple hashtags
	Use of multiple mentions	Whether the tweet contained multiple mentions
	Use of video	Whether the tweet contained video
	Use of imagery	Whether the tweet used imagery
	Use of URL	Whether the tweet included URL
	Use of emoji/s	Whether the tweet used emoji/s

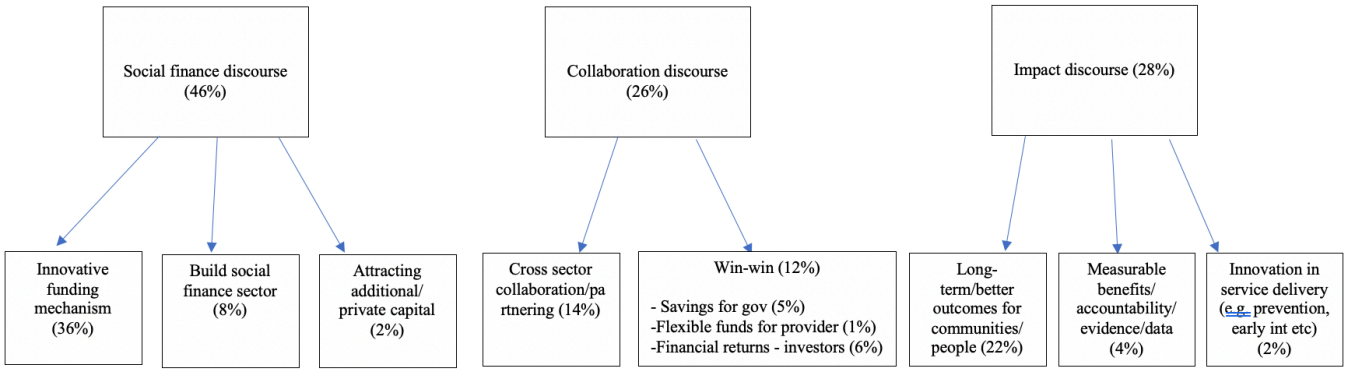
Discourses framework – purposive sample

After analyzing the random sample, I constructed a purposive sample (consisting of the most highly retweeted tweets). I applied an existing framework developed in a SIBs-focused discourse analysis study (Ormiston et al., 2020) (Figure 8.1). This method advanced my research agenda by allowing for an understanding of the extent to which discourses unrelated to evidence and primarily focused on the novelty factor, have been prioritized in the online discourse, potentially providing further evidence of institutional motives. For pragmatic reasons, I restricted the dataset to tweets that had been tweeted at least 5 times or more (283 qualifying tweets accounting for 2,211 tweets) and manually analyzed the discourses contained in these. Focusing on retweets reduced the risk that the findings apply to only a small group, as each individual tweet has been shared multiple times. Previous studies have used retweets to understand public attitudes, engagement with an issue or social movements (e.g. McNeil et al., 2012; Thelwall & Thelwall, 2020; Wang & Zhuang, 2017). There are many possible reasons for retweeting, although one reason (of relevance to this study) is to persuade others (Metaxas et al., 2015; Neiger et al., 2013). Thus, this section of the thesis also gave some insights into how rhetoric, rather than evidence, has been relied upon in the most highly reshared tweets.

When coding the retweets according to the pre-defined framework, I focused on a thematic analysis of the material rather than focusing on micro factors such as word order and propositional structures. While I used Ormiston et al's (2020) existing framework as the

starting point (Figure 8.1), I built on it by identifying several additional (largely cautionary) categories. After applying the framework, I conducted some additional coding, which involved identifying: 1) the types of 'events' mentioned; 2) the stakeholders who had retweeted; and 3) the evaluations and academic papers that were mentioned in retweets. This allowed for a deeper analysis of factors such as how negative and/or inconclusive results have been treated by members of the reform community, online. This is important to analyze because if stakeholders are pursuing SIBs for efficiency reasons, one would expect concern to be expressed when negative results have emerged (e.g. retweeting negative findings, drop off in interest, mentions of 'it does not work' etc). Conversely, if negative and/or inconclusive findings are drowned out in the retweets, this could provide evidence that stakeholders are pursuing SIBs for legitimacy reasons (and thus exhibiting low concern about the negative results).

Figure 8.1: Coding Framework – SIB Discourses



Source: Ormiston et al., (2020)

Additional analysis - measuring COI

In addition to applying the coding frameworks outlined above, I created a scale that measures the extent to which each tweeter has a COI. I considered how this had influenced factors such as sentiments expressed. The COI scale was applied to the random sample of Twitter data. This was an essential component, given my aim of understanding the role that SIB actors have played in influencing the diffusion of SIBs over time. I am particularly interested in the role played by stakeholders in the "fashion setting network" who have a vested interest in the success of SIBs. COI can arise in various ways in evaluation studies and can relate to possible financial or non-financial benefits (Eisner & Humphreys, 2012; Eisner et al., 2015; Resnik, 2000). Indirect benefits can be challenging to operationalize; thus, some instruments are limited to measuring financial COI (Eisner & Humphreys, 2012; Resnik, 2000). There are currently no instruments that have been developed for the SIBs context, despite the extent to which particular actors can benefit from the success of the idea.

As with past studies, I proposed an option that involves collecting information related to identifying a potential financial COI (Eisner & Humphreys, 2012; Krimsky et al., 1998; Friedman et al., 2004). I assumed that a COI is more likely if: 1) the tweet author (organization or individual) has been involved with at least one SIB contract; and 2) if the category of involvement was one where there could be a financial gain (e.g. investor, intermediary etc). Table 8.2 outlines the approach used; it was assumed that the more conditions that are fulfilled, the more likely it is that the tweeter can gain from the success of SIBs (and thus may be biased). While risk transfer (away from the commissioner) is a theorized benefit of SIBs, *"the government continues to bear a reputation risk, in case the service does not reach its desired outcome"* (de Mariz & Savoia, 2018, p. 308). Thus,

the COI was still considered medium for those involved as a commissioner or service provider.

Table 8.2: Scale for COI Levels

COI Level	*Factors associated with this*
High conflict	Organization or individual has been involved with one or more SIB projects; involved in an intermediary or investor capacity
Medium conflict	Organization or individual has been involved with one or more SIB projects; involved in a commissioner or service provider capacity
Low conflict	Organization or individual has been involved with one or more SIB projects as an evaluator
No conflict (broadly involved)	Organization who has not been involved in a SIB project at all but has had some broad (indirect) involvement – e.g. by publishing on SIBs
No conflict (completely uninvolved)	Organization who has not been involved in a SIB project at all and who has not had any identifiable indirect involvement

8.4 Ethical considerations

Twitter data presents various ethical limitations, particularly when incorporating qualitative methods. Studies show a disconnect between current research practices and users' views regarding the appropriate use of their data (Eurobarometer Survey, 2011; Williams et al., 2017). In Chapter 4, I outlined the ethical framework that has been applied to this study. As the use of direct quotes could make Twitter account holders identifiable, I tweeted each relevant account holder to inform them of the study and to give them the opportunity to opt-out. This approach has been applied in many other studies (e.g. Anderson et al., 2017; Talbot et al., 2020). It is consistent with the Oxford Central University Research Ethics Committee (CUREC) guidance and broader research on account holders' attitudes towards using social media data in research (Fiesler & Proferes, 2018). In total, nine account holders responded asking for more information about the study; only 1 account holder opted out of the analysis.

8.5 Findings

R1: What persuasive techniques have been used by involved stakeholders to frame the SIB funding model as a 'legitimate' reform to pursue over time?

To understand why the SIB concept spread from 2010 onwards, it is helpful to consider the various features identified by Hood and Jackson (1991) as being crucial to the 'acceptance factor'. The strategies of persuasion that were operationalized into the coding framework are presented in Table 8.3 with illustrative examples. Below the table, I present some descriptive statistics. While this chapter is primarily qualitative, the inclusion of some statistical observations (such as mentions of evidence, mentions of SIB alternatives and sentiments) allowed for the generation of insights which go beyond simply identifying the usefulness of Hood and Jackson's (1991) framework.

Table 8.3: Illustrative Quotes – Strategies of Persuasion used in Tweets

Feature	Description	Exemplar Tweet	Quotation/s from any links, or articles connected to tweet
Symmetry	Development of solutions that are seen to be 'exactly symmetrical' to the social problems experienced by the relevant audience (Hood and Jackson 1991).	*"In order to raise money for new initiatives in well-being, Flanders wants to issue social impact bonds" - Social investment can bring you money https://tijd.be/r/t/1/id/9982576?freeAccessToken=17afe710-8c60-465f-8871-7a4aa4ef9f13… via @tijd #socent #socinv #socialimpactbond"* *Note: the above tweet was translated to English using Twitter's "translate tweet" function.*	*"Innovative projects in the welfare sector often do not come about because the government's budget is limited and they do not want to take too much risk. That is why Flanders wants to launch social impact bonds following a foreign example".*
Metaphor	Persuasion can be achieved by the correct choice of metaphor, as successful institutions should be cognitive paradigms that appeal to both reason and nature (Douglas 1987).	*"New weapon available for governments #SocialImpactBonds cut costs & get #value4money pay for results only #DisruptAid"*	No links/attachments
Ambiguity	To be successful, the idea must be communicated to different audiences in a way that appeals to differing views and interests (Pocock 1971).	*"#socialimpactbonds should show govt a new funding option that shows what works. - #NHPC16"*	No links/attachments

Public Good & Private Interests	Private benefit must be justified in terms of public good rather than merely communicating the benefits to the individual or group of individuals promoting the idea.	*"Want to learn how #SocialImpactBonds can mobilize private capital to advance social good? visit http://ow.ly/8Xg9W"*	No links/attachments
Selectivity in Argument	A successful rhetorical argument must be highly selective in its choice of maxims to fit the desired conclusion. For example, arguments that contradict the desired conclusion may need to be omitted or downplayed.	*"Big news! #SocialImpactBond shows success! Read more, #PreK & #PayForSuccess change the odds. http://bit.ly/1LhFOSG @GoldmanSachs #SIB"*	*"Private funding allowed about 600 students to enroll in public and private preschool programs in 2013. Of those students, 110 4-year-olds were expected to need special education during their kindergarten year. But only one of the students — who are now in the first grade — has required special education, which translates to about $281,000 in cost avoidance for Utah's public education system".* *(Note: this downplays/completely omits evaluation issues, which later emerged)*
Suspension of Disbelief – intentionally framing an idea as 'new'	The final link to acceptance is the ability to suspend well-founded disbelief. One method SIB advocates use to suspend disbelief is presenting the idea as new to avoid it being dismissed as an idea that has already been tested and found unsatisfactory.	*"Very exciting #socialimpactbond project announced in NYC to fight recidivism: http://nyti.ms/N123hT"*	*"New York City, embracing an experimental mechanism for financing social services that has excited and worried government reformers around the world, will allow Goldman Sachs to invest nearly $10 million in a jail program, with the pledge that the financial services giant would profit if the program succeeded in significantly reducing recidivism rates.* *The city will be the first in the United States to test "social impact bonds", also called pay for success bonds, which are an effort to find new ways to finance initiatives that might save governments money over the long-term".* *"In an interview, Alicia Glen, the head of Goldman Sachs Urban Investment Group, said the company was confident that the program would work: "This is a new approach - no city has ever done anything like this before."*

Suspension of Disbelief - character assassination (attacking the doubters)	Additionally, character assassination can be used to suspend disbelief; this means "attacking the doubters as people rather than dealing centrally with their arguments".	*Verbatim tweet could not be provided due to inactive tweet; broadly, the tweet announced a new blog post.*	*"Sadly some tediously linear media negativists see only bad news in Goldman Sachs being paid a return (they have also taken a bath on the Rikers Island bond but the same media recidivists view that as "hah! they can afford it those ugly 1 percenters!").* *"Make no mistake, I may be a financial specialist and free markets / investment advocate but as previously stated: I clearly can see Goldmans is both a toxic brand for many who see it as a symbol of Wall Street's excesses (justifiably enough) BUT it is also the best advertisement for the SIB process as GS does have some great brains amidst its staff roster.* *If the brains trust of Goldman Sachs reckon SIBs work, then the process clearly has enormous potential".*
Suspension of disbelief - time pressure	Another common strategy for suspending disbelief is to appeal to the urgency of time, thus creating the perception that vital opportunities will be lost if immediate action is not taken. This may prematurely end the debate and prevent any objections calling for "piecemeal experimentation or controlled experiment before wholesale adoption".	*Verbatim tweet could not be provided due to inactive account; broadly, the tweet reflected on whether or not foundations should engage with PbR models such as SIBs.*	*"Should foundations support social enterprises and charities to participate in this market, particularly to serve more specific or harder to reach groups? My sense would be yes. If we don't act now then we are in danger of losing significant areas of charitable activity to the private sector, or worse ceasing to happen at all. I don't have an issue with the private sector per se, but this feels like a market where a mixed economy between private and charitable providers would have more chance of producing the best outcomes for people out of work".*

Source of framework: Hood and Jackson (1991)

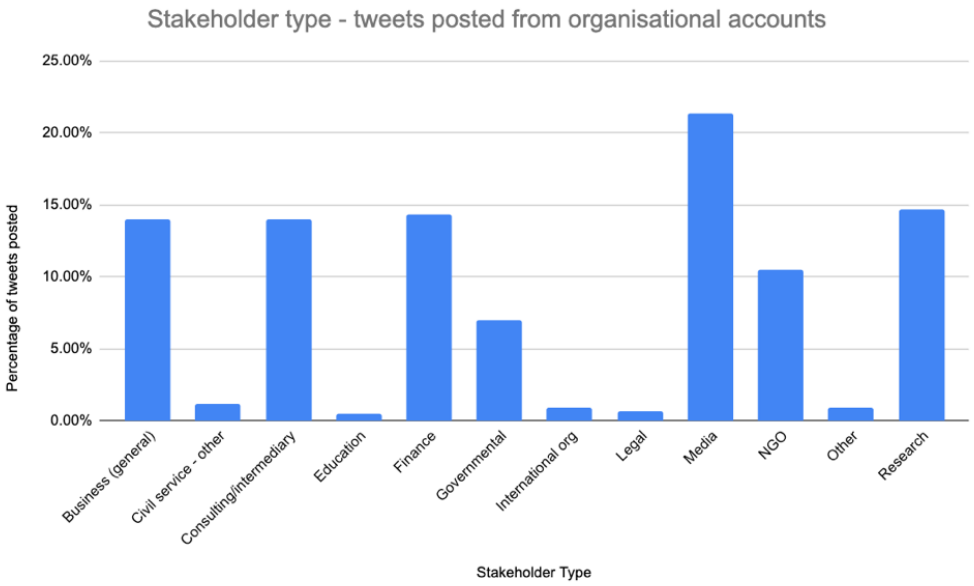
Table 8.4 provides a table of descriptive statistics for the random sample (742 tweets), which was used to explore the first hypotheses. It indicates the variability expected in the bar charts; for instance, it shows the extent to which there is an imbalance between particular stakeholder groups (e.g. 'directly involved' versus 'uninvolved' stakeholders).

Table 8.4: Descriptive Statistics – Random Sample

Statistic	*Value*
Number of tweets (total)	742
Number of English tweets	644
Number of non-English tweets	98
Number of tweets posted by organization account	428
Number of tweets posted by individual account	314
Number of tweets posted by directly involved stakeholders	238
Number of of tweets posted by broadly involved stakeholders	249
Number of tweets posted by uninvolved stakeholders	255
Total tweets posted by stakeholder with high COI level	121
Total tweets posted by stakeholder with medium COI level	60
Total tweets posted by stakeholder with low COI level	57
Total tweets posted by stakeholder with no conflict – broadly involved	249
Total tweets posted by stakeholder with no conflict –uninvolved	255

It was identified that 57.7% of the tweets contained in the random sample were posted by organizational Twitter accounts, such as the United Nations Development Programme (UNDP), and individuals posted 42.3% of the tweets. Regarding the tweets posted from organizational accounts, the most common category was media (21.3%), followed by research (14.7%), finance (14.3%), business (general) (14%), and consulting/intermediary (14%) (Figure 8.2). While international organizations (e.g. UNDP) have been influential in terms of retweeting (as discussed in part 2 of the 'findings' of this chapter), they are under-represented in the random sample (only 0.9% of tweets posted by international organizations).

Figure 8.2: Organizational Stakeholder Types



For tweets posted from individual accounts, Figure 8.3 shows that the most common stakeholder type was consulting (20.7%), followed by research (15.9%), and finance (13.1%). While SIBs have attracted substantial political attention, only 1% of the tweets posted by individual accounts were from politicians. Many 'individual' accounts (9.9%) could not be assigned to a specific category due to a lack of information in the relevant account profile.

Figure 8.3: Individual Stakeholder Types

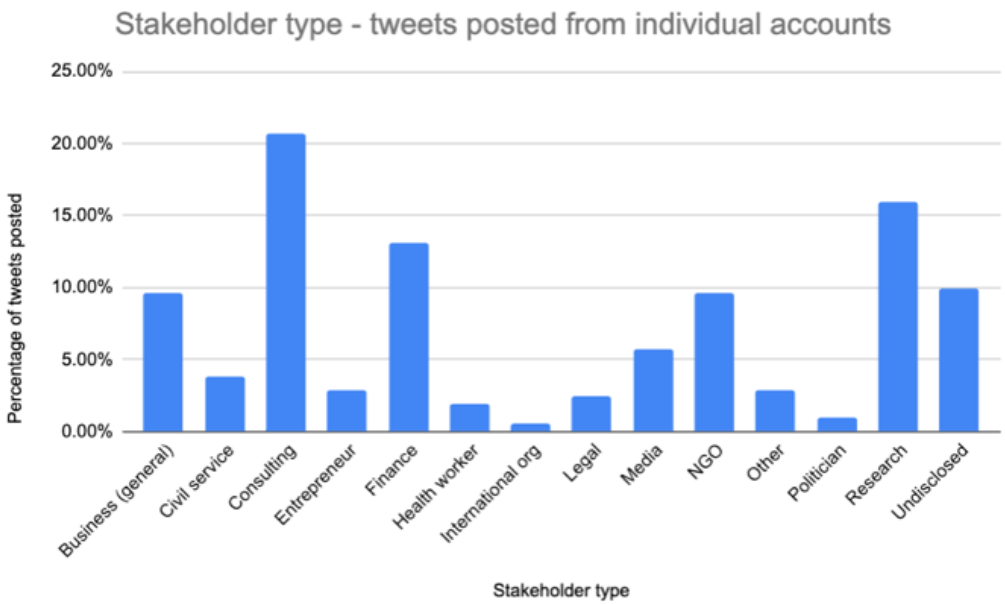
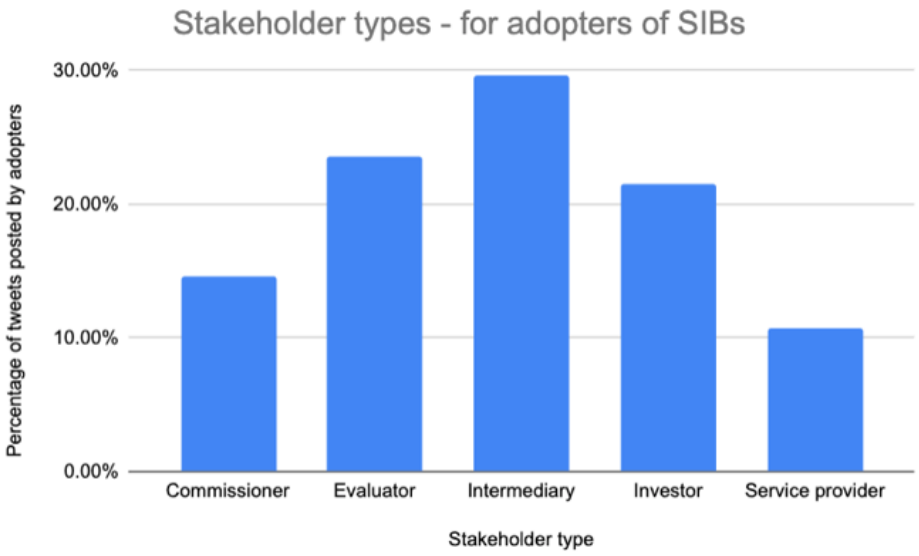


Figure 8.4 shows the breakdown of different stakeholder types for tweets posted by organizations (or individuals linked to organizations) featured in the GO Lab's INDIGO dataset (and thus verified as being directly involved in one or more SIB projects). It shows that some directly involved stakeholders were better represented than others. Intermediaries were the most prominent stakeholder type, followed by evaluators, investors, commissioners, and service providers. The percentage of evaluators was higher than expected, with service providers featuring to a lesser extent than anticipated.

Figure 8.4: Stakeholder Types for Adopters



Having described the demographics of the tweeters, I now present a series of insights (Figures 8 – 14) that assist with contextualizing the qualitative analysis which will follow. Firstly, Figure 8.5 shows that each persuasive technique included in the coding framework was quite prominent in the dataset, featuring in approximately a quarter of the tweets. Within the 'suspension of disbelief' category, 'framing the idea as new', was the most common technique; 'character assassination' and 'linking to national economic performance' rarely featured. Contrary to the original hypothesis, stakeholders with direct SIBs involvement appeared to use persuasive techniques to a lesser extent than uninvolved stakeholders.

Figure 8.5: Persuasive Techniques Used by Type of Stakeholder[7]

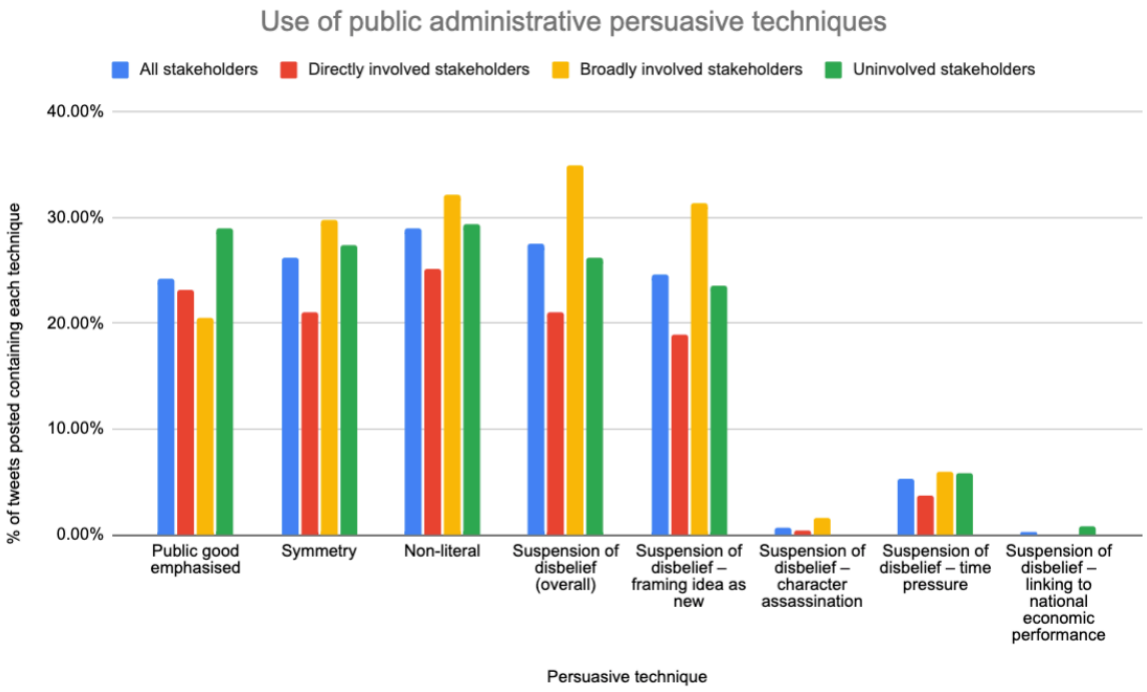
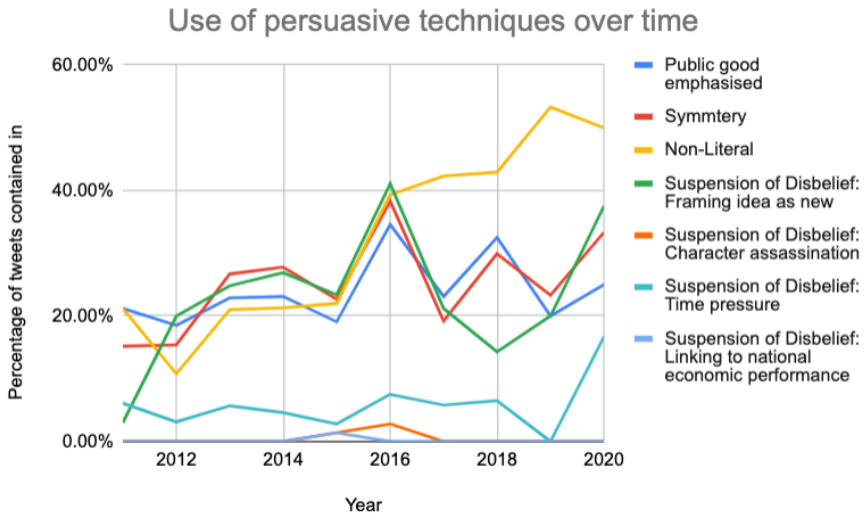


Figure 8.6 shows that the use of the persuasive techniques differed over time, with many techniques (e.g. symmetry, time pressure, emphasizing public good, framing as new idea) growing in recent times. A particularly unique insight is that the idea continues to be framed as 'new', even though the first SIB contract was launched over twelve years ago.

[7]Each group used the persuasive techniques to a different extent; in many instances, a single tweet contained multiple persuasive techniques, meaning that the percentages do not tally up to 100%.

Figure 8.6: Use of Persuasive Techniques Over Time



In addition to using public administration persuasive techniques, some Twitter-specific techniques were also identified in the dataset. Figure 8.7 shows that directly involved stakeholders used many of these techniques to a slightly greater extent. The most common techniques were the use of a URL and multiple hashtags; videos and emojis rarely featured. Despite this, the difference appears to be minimal, and thus, it may be overinterpreting the data to claim a meaningful difference between directly involved and uninvolved stakeholders.

Figure 8.7: Use of Twitter Persuasive Techniques

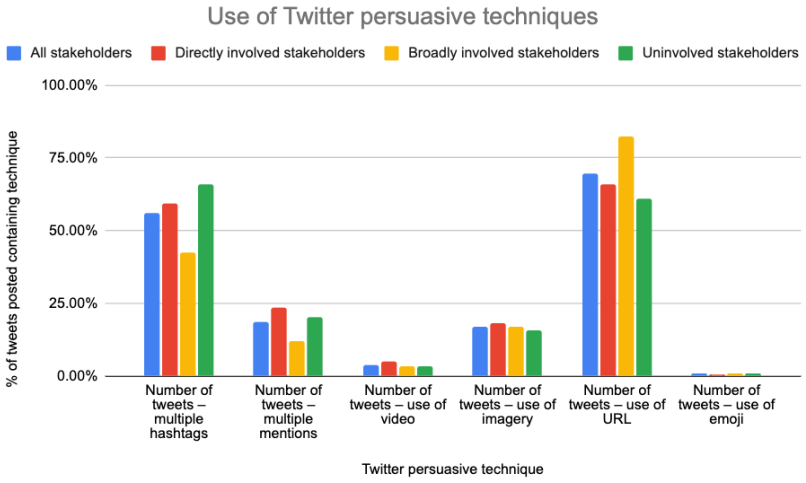
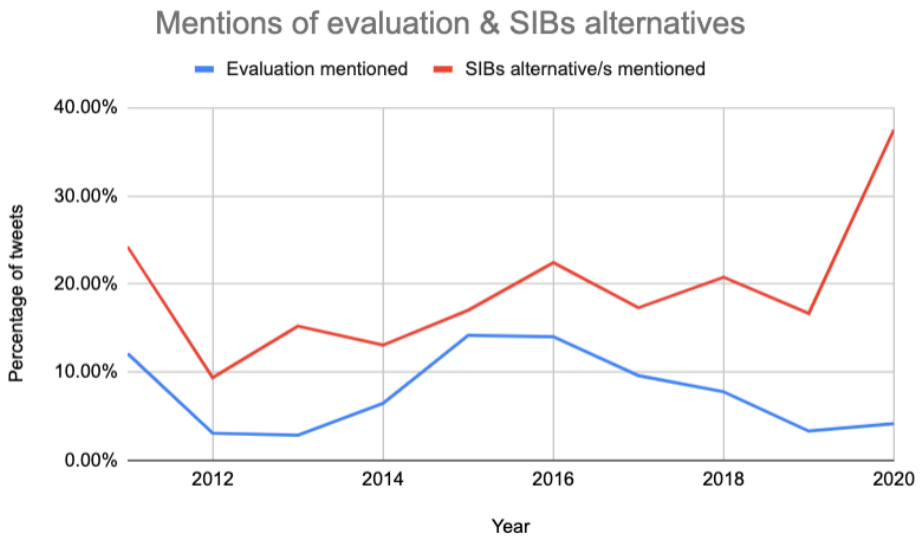


Figure 8.8 shows that the mention of specific evaluations peaked in 2015-2016 but has declined since then. Thus, evidence has been referred to less over time, even though more SIB projects have now been completed and more evaluations have been published. At the same time, mentions of SIB alternatives have increased in recent times, suggesting that interest in the innovation is declining and that it has not been institutionalized (Zucker, 1983).

Figure 8.8: Mentions of Evaluation and Alternatives – Over Time



As shown in Figure 8.9, involved stakeholders have mentioned EBP to a greater extent; and mentioned SIB alternatives less, on average. This could be reflective of SIB stakeholders being satisfied with the results of SIBs (and having concrete data to support the benefits). Conversely, it could be linked to stakeholders wanting to persuade other stakeholders to adopt the innovation, potentially for their own self-interests. Concerningly, mentions of evaluation are low for both stakeholder types and overall.

Figure 8.9: Mentions of Evaluation and Alternatives – by Stakeholder Type

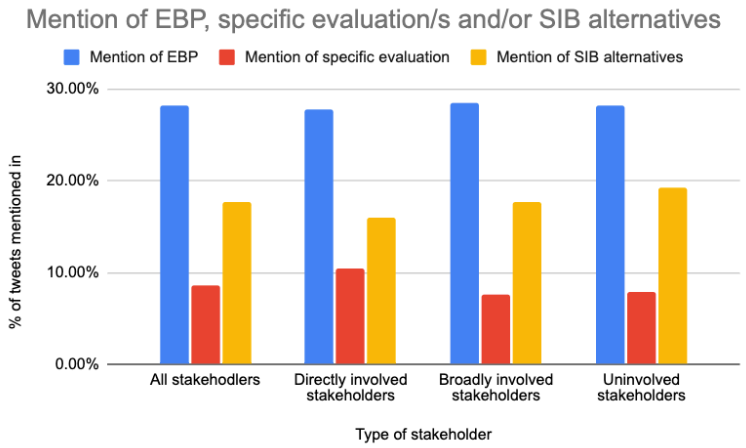
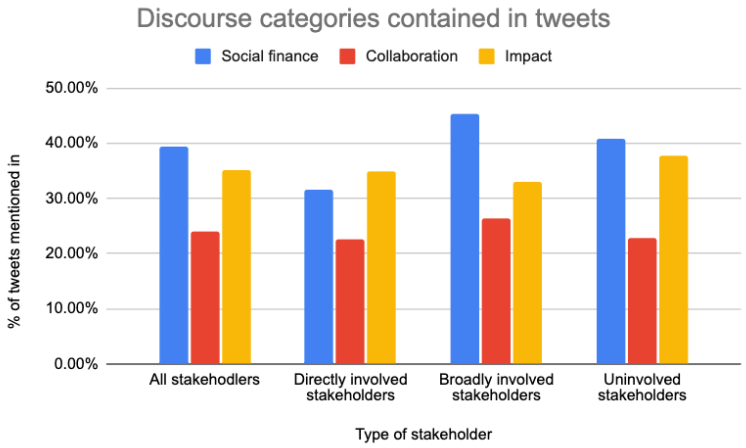


Figure 8.10 demonstrates that involved and uninvolved stakeholders mentioned the 'collaboration' and 'impact' discourses to a similar extent; however, broadly involved stakeholders more frequently mentioned the 'social finance' discourse (focused on the role of the private investor, innovative funding model and market mechanisms). As with the discourse analysis (discussed later in this chapter), all stakeholders used the 'collaboration' discourse category to a lesser extent.

Figure 8.10: Discourses Mentioned – by Stakeholder Type



As depicted in Figure 8.11, the use of the 'social finance' and 'impact' discourses has been increasing recently, while mentions of the collaboration discourse have been decreasing.

Figure 8.11: Discourses Mentioned – Over Time

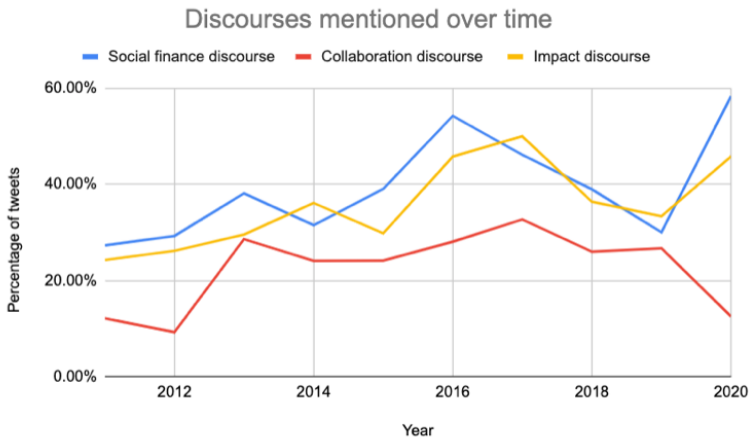
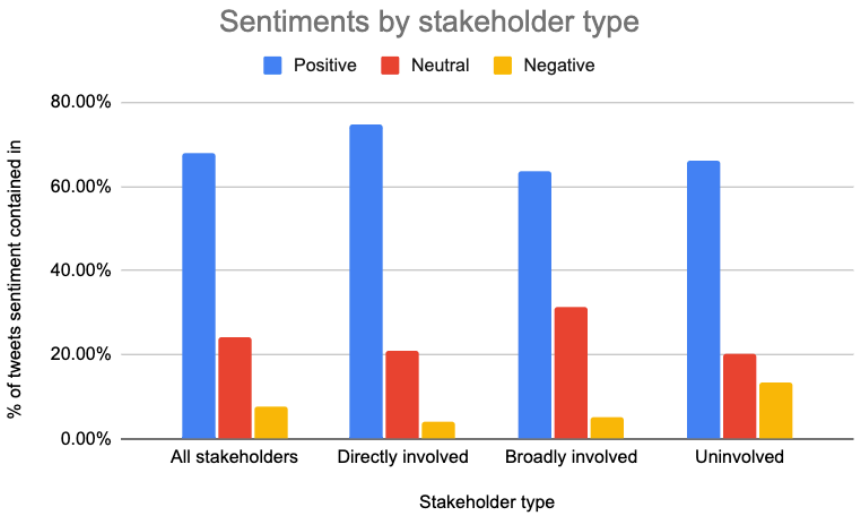


Figure 8.12 provides further evidence (consistent with Chapter 7) of positive sentiments being over-represented in the dataset, with far fewer negative sentiments. Directly involved stakeholders had a slightly higher percentage of positive and fewer negative sentiments than uninvolved stakeholders. The sentiment scores contained in this chapter are more nuanced than those calculated by the pre-built packages (Chapter 7), as the scores calculated by pre-built packages only considered the wording of the tweet (and struggle to interpret humour, emojis etc), while I traced the entire tweet, including URLs and videos.

Figure 8.12: Sentiments by Stakeholder Type (Whole Time Period)



Consistent with the Chapter 7 findings, Figure 8.13 shows that positive sentiments are far more prevalent in the data, with fewer negative tweets. As with the Chapter 7 findings, positive sentiments dipped but have slightly risen again in recent times. Contrary to Chapter 7, however, negative findings have also been gradually increasing, potentially as inconclusive (and sometimes negative) findings have accrued in the literature. In comparing the sentiments generated using pre-built packages (Chapter 7) to those that I generated from manual coding (random sample only), I found a fairly low proportion of matching sentiments (RSentiment: 52.29%; SentimentR: 48.92%; Rsyuzhet: 48.65%; SentimentAnalysis: 45.28%; and MeanR: 37.60%). It is worth noting that the percentage of tweets with matching sentiments (i.e. match between automated and hand-coded sentiments) was slightly higher when I coded sentiments for the entire dataset (Chapter 7), although notable differences still remained (Rsyuzhet: 56.98%; SentimentR: 56.74%; Sentiment Analysis: 56.63%; RSentiment: 55.96%; and MeanR: 51.97%). The extent of the differing sentiment categorizations (between the manual and automated coding), is presumably due to the inherent difficulties associated with coding sentiments in short tweets (e.g. difficult to interpret sarcasm, humour etc), along with the fact that I coded sentiments towards SIBs specifically, while the automated packages coded the

sentiment of the entire tweet (which or may not have been reflective of the tweet author's attitude regarding SIBs). Also, the automatically generated sentiments only considered the content of the tweet, while I analyzed all included links, videos and images, for the tweets contained in the random sample. When coding sentiments for the entire dataset, (for Chapter 7), I analyzed the tweet content only, for pragmatic reasons.

Figure 8.13: Sentiments Over Time

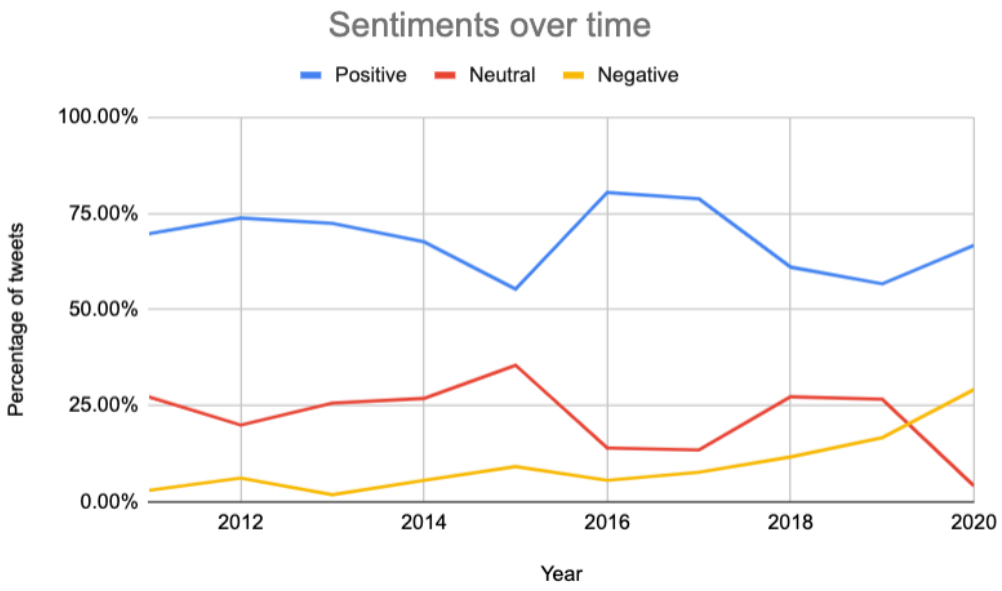
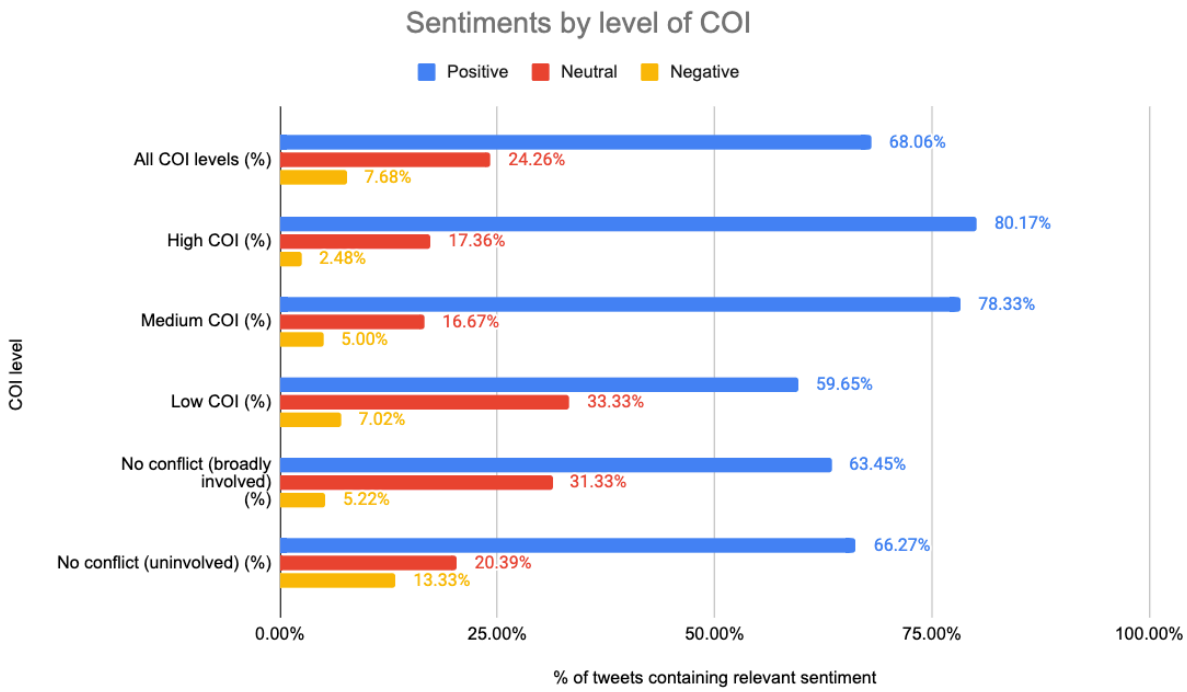


Figure 8.14 shows that the COI level appeared to influence the sentiments expressed. On average, those with a high COI expressed more positive and fewer negative sentiments than those with a medium, low or no COI. There was a noticeable difference between the sentiments in tweets posted by those with a medium COI and those with a low COI (fewer positive and more negative, on average), thus demonstrating the importance of separating different types of stakeholders in the COI scale. Stakeholders with no COI (broadly involved) posted a similar percentage of positive tweets compared to stakeholders with no COI (uninvolved); however, those with no COI (uninvolved) posted more negative tweets, on average, as expected.

Figure 8.14: Sentiment Category by COI Level



Having outlined some general observations relating to the random sample, each persuasive technique introduced earlier in this chapter is discussed in further detail below.

Symmetry

SIBs were often presented as a solution which is 'symmetrical' to the problems experienced by stakeholders in several ways. Firstly, many tweets highlighted the problem of a lack of data and evidence; and noted that SIBs are a solution as they can *"put evidence and outcome measurement at the heart of contracting".* For example, an article linked to a tweet noted that *"the process of paying a service provider based on outcomes pushes the government to introduce data monitoring and evaluation systems above and beyond the ones it may be using today"* (Belinsky, 2012). Similarly, various tweets noted that there had been increasing expectations and demands on politicians for more results-driven investments and

enhanced accountability, thus leading to increased interest in funding mechanisms that can transfer financial and political risk to third parties.

Secondly, SIBs have been presented as a solution to fiscal constraints, particularly since the Global Financial Crisis (GFC), which resulted in a decline in the public sector's capacity to sustain its welfare system. In one tweet, an infographic was included, which noted that while city, state and federal budgets are declining, *"the social challenges those governments face aren't going away".* Likewise, it was noted that citizens continue to expect more effective services despite shrinking budgets. Policymakers are thus turning to SIBs to 'fill this gap'. An article linked to a tweet noted that in France, organizations have 'suffered' from the effects of budgetary restrictions since 2010 and amid this 'crisis', "*it is legitimate for the public authorities to seek new funding levers"* such as SIBs (Pigneur, 2016).

A related issue highlighted in the tweets (and related to the above budgetary issue) was the lack of money for preventative services, with 'cash-strapped governments' often unable to allocate sufficient funds to early intervention. For instance, an article noted that the 'sad reality' in Denver is that the jail *"has become one of Colorado's largest providers of mental health services"* (Parvensky & Clark, 2018). The article proposed SIBs as a 'homegrown solution' to the problem of underfunded preventative services. It emphasized the importance of this, noting that *"the systems in place to deal with behavioral health challenges are often under-resourced, resulting in the last resorts of jails, emergency rooms, and homeless shelters across the country being filled with individuals who have unmet behavioral and mental health needs"* (Parvensky & Clark, 2018).

Additionally, many tweets demonstrated that SIBs fit the desire for the appearance of something new, which differs from past reforms. For instance, one article linked to a tweet highlighted that traditional forms of financing, such as donations and grants, do not allow social entrepreneurs to innovate and grow, thus leading to the need for new ways of financing, such as SIBs.

> *"Many of us have grant-funded charities and projects in the past to help fill the void in resettlement support for prisoners serving sentences of less than 12 months. In the SIB we saw an opportunity to demonstrate on a much more ambitious scale that support for these prisoners not only reduces crime and the numbers of future victims, it makes financial sense as well"* (O'Kelly, 2014).

In addition to the above examples of symmetry, it was noted that investors increasingly want to invest responsibly. One article, linked to a tweet, stated that *"more modern investors want to allocate their money in alignment with their hearts" and that SIBs provide an "exciting new way [for them] to do so"* (Lomax, 2015)

Similarly, SIBs have been viewed as a way to address the growing need for corporate giving, as highlighted in this tweet: *"#innovative #financial models such as #socialimpactbonds by Goldman Sachs, are being used to address the need for corporate giving. #smg14".* A related point was the need for some stakeholders to enhance their image, with SIBs being proposed as a solution for reputational issues. Specifically, it was noted in one article, linked to a tweet, that *"Goldman suffers from a dubious image"* and that an investment in SIBs contributes to *"a good, or at least slightly better, name of the investment bank"* (Kooiman, 2012).

Metaphor

Metaphors are also considered a persuasive technique that can increase an idea's acceptance. Various metaphors were identified in the SIBs Twitter dataset. For example, one article linked to a tweet referred to 'cash-strapped' governments' who are 'tightening their belts' and thus needing to resort to SIBs to achieve more with less.

> *"As cash-strapped governments continue to tighten their belts, 'social impact investing' – in which the wealthy fund schemes or projects that benefit society but also provide a return – is becoming more and more appealing"* (Caplen, 2016).

Another noticeable metaphor was the use of the word 'weapon' in the following tweet, likely as a way to emphasize that the idea can be deployed to gain an advantage or to defend oneself, thus overcoming organizational problems.

> *"New weapon available for governments #SocialImpactBonds cut costs & get #value4money pay for results only #DisruptAid".*

Likewise, an article connected to a tweet used the word 'shackled' in relation to Non-Governmental Organizations (NGOs), thus using imagery to emphasize the shortcomings of the current system and to imply that without SIBs, NGOs are confined and prevented from taking specific actions. The article emphasized various benefits, such as unlocking impact data and enhancing transparency, noting that *"for a nonprofit sector that to date has been shackled with an opaque and inefficient capital market, any of these benefits would be transformative"* (Beck et al., 2016).

Another strong example of a metaphor appeared in the following tweet referring to SIBs as the 'tip of the iceberg' in the trend of performance-based contracting:

"#socialimpactbonds may go nowhere but are tip of iceberg in emerging trend of performance-based contracting @SIRCtweets #smartcitiesweekdc".

Ambiguity

Part of the appeal of SIBs can also be traced to the ambiguity used when promoting the idea. Through the use of broad and vague terms, SIBs have appealed to different groups of stakeholders for different reasons. For instance, vague and multi-faceted concepts such as 'performance' were regularly used in the tweets, potentially carrying a wide variety of meanings. The concept of 'performance' or 'performance-based' can potentially refer to effectiveness, efficiency, quality, availability, accessibility, consumer satisfaction, or any combination of these concepts. Thus, the notion of 'improving performance' could relate to improved financial performance and/or improved service quality. Additionally, the broad concept of investing in 'what works' featured explicitly in multiple tweets. For instance, one tweet noted that *"#socialimpactbonds should show govt a new funding option that shows what works".* Another tweet emphasized that *"#socialimpactbonds improve outcomes increase #innovation rate & learning of what works #rcsocialimpact".* Finally, another tweet noted that *"organizations are excited about the flexibility they get to do what works"" from #socialimpactbonds".*

An additional example of a concept often vaguely referred to in tweets is accountability. For instance, one tweet noted that Bloomberg Philanthropies are *"in #SocialImpactBonds not bc #impinv but bc model for public-private partnerships + accountability #whinnovation".* Another tweet emphasized that it is *"Great to see UK leading commissioning of #SocialImpactBonds These programmes create space for #Innovation #Flexibility & #Accountability enabling operational experts & participants to*

co-design models. This approach has meant @KBOP_Kirklees was able to become truly #AssetBased".

Emphasizing public good

Effective persuasion may be considered *"the art of converting self-interest into an issue of high principle"* (Gittins, 1988). Thus, Hood and Jackson (1991) note that persuasive ideas are often presented in terms of public good rather than private benefit. An example of this could be seen in tweets that commended stakeholders such as the Rockefeller Foundation for generating public good without acknowledging that they have a clear vested interest in the idea's success. For instance, one tweet failed to mention anything about private benefit, instead only noting that the funding provided by the Rockefeller Foundation would assist with streamlining services with better metrics and providing 'insurance' to the poor. Another tweet, posted by an individual with direct SIBs experience (intermediary capacity), focused on public good, highlighting that a *"New Report Argues that a #SocialImpactBond could Strengthen Workforce Development in Ghana http://t.co/DFSoUb1HJH #DevImpactBond #SIB #DIB".* Tweets focused solely on public good often emphasized the impact generated for particularly vulnerable cohorts. For instance, the following tweet noted that problem families would be assisted through a SIB *"#ILikeItWhen #SocialImpactBonds #PaymentByResults r used to help #ProblemFamilies Checkout @StGilesTrust @StGilesTrustPRS @StGilesTrustSOS".* Other tweets in this category focused on social outcomes generated, rather than investor repayments; for instance, one tweet noted that: *"The first #SocialImpactBond in the UK reduced re-offending rates by 8.4%. This and more good news in our report http://t.co/0BK8JQ3e3Y".*

Selectivity in the argument

Another element identified as being strongly linked to the 'acceptance' (or otherwise) of a reform idea is selectivity in the argument. This involves selectively using examples rather than relying on rich data. For example, in relation to New Public Management (NPM), Hood and Jackson (1991, p. 192) noted that *"its advocacy was rich in anecdotes about extravagant failure or waste in the 'old' public management".* Similarly, the Twitter dataset appeared to be dominated by cherry-picked 'success' stories which fit the prevailing argument as to why SIBs should be pursued; examples of SIB shortfalls and criticisms were far less prominent. A prominent example of this was the framing of the Utah PreK SIB as a success, despite strong criticisms of the project in the broader literature in relation to the evaluation design. The following tweet overlooked the serious evaluation issues raised, framing the project as a success: *"Big news! #SocialImpactBond shows success! Read more, #PreK & #PayForSuccess change the odds. http://bit.ly/1LhFOSG @GoldmanSachs #SIB".* Several additional tweets also framed the project as a success: *"Goldman Sachs received the first #SocialImpactBond #payoff in USA. Special education for #Utah children was avoided. https://t.co/uNI6AWfSEb".*

While some tweets (such as the one below) highlighted shortfalls and provided a critical perspective, these featured much less in the dataset and were largely drowned out by the success stories and general enthusiasm surrounding the idea.

"#socialimpactbonds project #NYCABLE project 4Incarcerated youth, didnt work, R these projects worth it? @paulalantz talks us through #NHPC16".

Suspension of disbelief

The ability to suspend disbelief about an idea is crucial, thus leading to the use of tactics such as character assassination, assertion of time pressure, and intentionally framing an idea as 'new' (Hood & Jackson, 1991). These sub-categories are each discussed below.

Character assassination

Although less prevalent in the dataset, there were some notable examples of character assassination. For example, in an article linked to a tweet, doubters of the SIB concept were referred to as *"tediously linear media negativists*" who view *"anybody being paid for taking a risk as a bad thing"* (Young, 2015). Thus, sceptics were portrayed as being overly pessimistic, reactionary and 'out of tune' with the times. Another article linked to a tweet labelled stakeholders who dismiss ideas related to privatization as 'dull reactionaries'. It noted that *"another dull reactionary discusses the evils of privatizing health care which only kills the debate about how we pay for better medicine to an aging population"* (Young, 2016). Likewise, an article linked to a tweet branded a political party that had decided not to adopt SIBs as the party that 'says no' and is fearful of social innovations.

Assertion of time pressure

Through building up a sense of crisis and urgency, SIB proponents have also been able to *"quash 'try before you buy' objections calling for piecemeal experimentation"* (Hood & Jackson 1991, p. 194). For instance, identifying a $30 trillion funding shortfall for the SDGs and no agreement as to how to cover it was regularly cited as a way to build urgency around the SIB idea. An article linked to the tweet below specifically noted that the funding gap *"cannot be bridged without private investors"* (Brown, 2018).

> *"Are the financial community well positioned to deliver #socialimpact? Gordon Brown makes the case for #socialimpactbonds in the @FinancialTimes #impactinvesting #golabnews @OfficeGSBrown https://ft.com/content/57b7f408-b80b-11e8-a1d8-15c2dd1280ff".*

Some tweets and articles were even more explicit about the urgency; the article below (linked to a tweet) specifically noted that there is 'no time' to waste.

> *"Donors and development agencies are keen to make the most effective use of their budgets. They strive to generate sustainable impact on a much larger scale. And rightfully so: the new Sustainable Development Goals are literally calling for it, as the World Economic Forum outlines in its 2015 guide to blended finance. And there is no time to be wasted"* (Struewer & Moehrle, 2015).

In some instances, tweets and associated articles focused on the economic and social development challenges faced in a specific location, emphasizing the persistence and enormity of the issues. For example, an article linked to a tweet highlighted the urgent need for a new funding approach in France, noting that:

> *"The figures are implacable: France is world champion in public spending, long-term unemployment has never been so high, extreme poverty increased by 30% from 2002 to 2011, the number of households receiving RSA is increased from 1.8 million at the end of 2010 to 2.4 million at the end of 2014"* (Enterprise Institute, 2016).

In another tweet, a video was provided, which emphasized the problem of lead poisoning in USA, calling for urgent attention and flagging that SIBs could play a key role in overcoming this issue. Several articles published since this tweet have emphasized the extent of this problem (e.g. Hauptman et al., 2021).

Urgency was further conveyed in various tweets by referencing how the COVID pandemic has exacerbated injustices and inequities. For instance, one tweet linked to an article called for a *"reset not a recovery"*. The linked article emphasized that we cannot

simply return to ‘normal’ as the *“crisis has offered our world a chance to change for the better”* (Yousafzai, 2020).

Presenting an idea as new

Another common strategy of persuasion involves presenting an idea as new, to avoid any disbelief on the grounds of it having been tried already and shown to be unsatisfactory. Many tweets described SIBs as a 'new', 'novel' and/or 'innovative' type of funding tool. They used wording such as ‘experiment’, ‘experimental mechanism’ and ‘promising’ (i.e. unproven), despite the fact that *“in reality, of course ‘better public management’, through the adoption of business methods, is a doctrine which has a long and chequered history”* (Hood & Jackson 1991, p. 194).

An example of a SIB being framed as a completely new idea could be seen in an article linked to a tweet focused on the issue of homelessness in Santa Clara County. It noted that *“Such deals are complicated. And their approach is so new that only a handful of state and local governments have begun experimenting with them”* (Wiltz, 2016). Additionally, a tweet that introduced SIBs as *“a virtuous collaboration between public and private”* included a link to an article which emphasized that traditional funding models are reaching their limits. The article explicitly stated that *“not everything has been tried”* and proposed SIBs as a new solution, noting that *“all over the world, the social sector is turning to a new innovative model: Social Impact Bonds, or social impact contract”* (Enterprise Institute, 2016).

Similarly, an article linked to a tweet described SIBs as a *"groundbreaking"* concept which holds *“extraordinary potential”* (Beck et al., 2016). It noted that the concept *“brings together donors, impact investors, and non-profit organizations to fund impact in a*

completely new, performance-driven way" (Beck et al., 2016). Another tweet excitedly announced the launch of a New York City (NYC) SIB aimed at fighting recidivism and emphasized the new, experimental nature of the funding model. The tweet stated the following: *"Very exciting #socialimpactbond project announced in NYC to fight recidivism: http://nyti.ms/N123hT".* The associated article commended NYC for *"embracing an experimental mechanism for financing social services that has excited and worried reformers around the world"* (Chen, 2012). Specifically, the head of the Goldman Sachs Urban Investment Group noted that *"this is a new approach – no city has ever done anything like this before"* (Chen, 2012).

While many tweets (and associated articles) completely overlooked prior related reforms, some did provide important context. For example, one article linked to a tweet noted that SIBs fit within three (not mutually exclusive) frameworks, one of which is NPM (Painter et al., 2018). Concerning NPM, the article noted that SIBs are an extension of this approach, as the funding model incorporates markets, managers and measurement and aims to generate greater efficiency and accountability in the provision of public services (Painter et al., 2018).

Linking to national economic performance

This final sub-category was extremely under-represented in the dataset, although it did appear in a small number of tweets. It involves linking national economic performance with the quality of public management; for instance, the belief that efficient management is crucial for economic revival. One tweet explicitly noted that the use of "meaningful public-private partnerships" would help the economy: *"Help #economy Moving towards meaningful public-private partnerships @NJpoliticsprof @SetonHall #socialimpactbonds https://t.co/9w4Pe2np8V"*. Another tweet noted that the French economy is in decline and

that SIBs could assist with restoring confidence in French society and improving the quality of public missions.

R2: What discourses can be identified in the highly shared tweets; and to what extent have particular discourses dominated/others been drowned out?

The second research question focused on identifying the discourses contained in the highly shared tweets. The purposive sample consisted of 283 qualifying tweets (retweeted at least 5 times), accounting for 2,211 retweets. Some critical insights were gained from looking at the types of evidence (Table 8.5) and events (Table 8.6) mentioned in the retweets. Table 8.5 highlights that practitioner reports were the most common type of evidence mentioned (in 31 tweets, accounting for 198 retweets). This is notable given that these reports are often written by those with a vested interest in the success of SIBs. The retweets often spoke vaguely about the promises of SIBs in terms of enhancing EBP and accountability, yet very few retweets mentioned specific academic studies or evaluations.

On the one hand, this could be aligned with Hood and Jackson's (1991, p. 17) finding that persuasive administrative ideas often *"rest on soft data and soft logic"*, with stakeholders selectively drawing on examples and using vague terminology. Alternatively, it could be linked to the fact that as an idea spreads, later adopters are less likely to focus on how well the idea works; and proponents may 'dilute' their proposals to make an idea more broadly appealing (Best, 2006). Finally, it could suggest a level of 'decoupling' surrounding the use of evidence in this context, with policymakers potentially symbolically committing to focusing on outcomes (and tweeting about that) but failing to achieve this in practice.

Table 8.5: Types of Evidence Mentioned in Highly Retweeted Tweets

Publication Type	Number of top retweets mentioned in	Cumulative number
Practitioner report	31	198
Evaluation report	15	110
Journal article	9	54
Government report	5	33
Book	2	20

Table 8.6 outlines the categories of events most commonly mentioned, demonstrating an over-representation of events associated with the launch of a SIB (i.e. pre rather than post-SIB implementation). For instance, the launch of a SIB project was mentioned in 41 of the tweets (accounting for 321 retweets in total). In comparison, the announcement of results (including mention of an impact evaluation) only featured in 2 tweets (accounting for 15 retweets). It is essential to consider possible biases introduced by the fact that retweets are overly focused on future contracts, with few focusing on completed projects. As discussed earlier, there has historically been a pro-innovation bias in diffusion research, with the early literature emphasizing the relative advantage of early innovations and promoting rapid diffusion (Rogers, 1995, pp. 122-123). Table 8.6 suggests that the pro-innovation bias exists in the context of SIBs, potentially contributing to positive tweets about SIBs receiving more attention and being shared more widely.

Table 8.6: Types of Events Mentioned in Highly Retweeted Tweets

Category	Number of retweets mentioned in	Cumulative retweets
Launch of a SIB	41	321
Public event - conference, roundtable, workshop etc	44	303
Grey literature released	39	265
Announcements about other enabling factors (e.g. passing of legislation)	15	88
Results announced (all)	12	86
Announcements regarding SIBs videos, documentary etc	15	85
Launch of SIBs Fund or SIBs Fund update	9	70

Academic research released	7	40
Meetings between stakeholders re SIBs	6	38
Results announced (impact evaluations only)	2	15
Public awards	2	10
Other procurement updates	2	10

Table 8.7 provides exemplar quotations for each of the dominant discourses identified. It includes the three major discourses already presented in the SIBs literature ('social finance' discourse; 'collaboration' discourse; and 'impact' discourse) (Ormiston et al., 2020) and the additional categories and sub-categories identified in this study. The 10 underlying discourses identified by Ormiston et al., (2020) were very transferable to the online Twitter discourse, with strong examples of each of these identified. Regardless, there was a clear need to add further discourse categories, focused on the possibility (and in some cases evidence) of harm and the excessive hype surrounding the idea.

Table 8.7: Exemplar Quotations

Discourse	***Exemplar quotation***
Social finance discourse	
Innovative funding mechanism	*"The state of Connecticut launched its first #SocialImpactBond, an innovative and #sustainable financial tool. https://t.co/4MDduVbWWD"*
Build social finance sector	*"Great post at @CGDev on our new facility for testing #SocialImpactBonds in #LAC http://t.co/WUukB2NZXk #SIB"*
Attract additional/private capital	*"A momentous start to 2018 as global social impact bonds surpass 100, with $392M capital raised". #108SIBsForImpact #socialimpactbonds #sibs @socialfinanceIL @SocialFinanceUS https://t.co/Mh5KeSF89M"*
Collaboration discourse	
Cross sector collaboration/partnering	*"To give young people leaving care more support, we're working together & thinking outside the box! Kicked off Australia's largest #SocialImpactBond with @timpallas* *@anglicarevic* *@vincentcarevic at #compasslaunch https://premier.vic.gov.au/making-things-fairer-for-young-people-leaving-care/ #springst"*

Win-win-win	*"#SocialImpactBonds a "triple-win scenario" of positive social outcomes, gvmt savings, and investor profit http://t.co/6s72NECeKX @MplsFedCMD"*
Impact discourse	
Long-term/better outcomes	*"One year on, the #Newpin #SocialBenefitBond has helped many families and generated positive returns to investors - http://t.co/MF4meGTVJ9"*
Measurable benefits/accountability/data	*"We evaluated Birmingham's StepDown programme:moving from #residentialcare to #fostercare #socialimpactbond https://t.co/v49Dfyhmwj"*
Innovation in service delivery	*"Fantastic piece on our locally run & innovatively designed #socialimpactbond project #SFHomelessProject"*
Potential harm discourse	
Financialization (marketizing vulnerable people, profit motives prioritized etc)	*"#PayForSuccess First Modeled=Utah Kindergarten Used #MTSS,#PBIS,#RTI 2Discourage Referring Special Ed Kids #IDEA Funds Used 2Train Reg #Teachers= #AppliedBehaviorAnalysis #SocialimpactBonds Used 2PAYBACK GoldmanSachs 2NOT Send Kids 2 SpEd So SpEd FundsUsed 2 Not Refer SpEd Kids https://t.co/rimgetXMVF"*
Unintended consequences	*"#SocialImpactBonds are not always the answer for social problems. Intervention without real knowledge about the root cause of a problem in a community can cause more negative consequences than benefits. #socialimpact https://ssir.org/articles/entry/the_downside_of_social_impact_bonds#"*
Problematic power dynamics	*"even poverty can be profitable" Look For #SocialImpactBonds #PayForSuccess Using #Data 2 Identify #TitleI Kids For Direct Interventions Toward #CommonCore & #SocialEmotionalBehavioral #Conditioning* *Social Impact Investing, Poverty, & Bill Gates' New Plan https://t.co/URhUaZJhOK"*
Excessive hype discourse	
Idea is not 'new' despite claims	*"#Socialimpactbonds are just #privatization in sheeps clothing. They have no place in our health care system https://t.co/vHEBgAt7it #cdnpoli"*
Benefits not eventuating in practice	*"How should we evaluate the success, or lack thereof, of the #socialimpactbond pilot program? @carolinefiennes http://t.co/zcYJ6w8L9y"*
Need to consider other models	*"There's more to social investment than social impact bonds, UK politicians are told: https://t.co/yH0PyXPJnG #SocialImpactBonds"*

While the 'social finance' and 'collaboration' discourses dominated press releases (Ormiston et al 2020), the Twitter dataset predominantly featured the 'social finance' discourse and the 'impact' discourse (see Figures 8.15 and 8.16), with 'collaboration' far less represented. The dominance of the 'social finance' discourse is expected, given that this discourse, which privileges the role of the private investor and market mechanisms, is aligned

with NPM. SIBs are perceived by many as a new type of arrangement under the broader NPM umbrella, involving the incentivization and privatization of public service provision through the development of quasi-markets (Hevenstone & von Bergen, 2020; Pollitt & Bouckaert., 2017). Further, within the 'social finance' discourse, there is a category focused on the innovative nature of the funding mechanism, which regularly featured in the tweets. Tweets containing positive sentiments often emphasized novelty, while negative tweets often noted that the SIB funding model is not a 'new innovation'. As the 'collaborative' discourse, which is more aligned with the shift to New Public Governance (NPG), featured less regularly than the other discourses, it appears that the other discourses were perceived as more exciting and/or more persuasive by SIB advocates (Joy & Shields, 2020; Vecchi & Casalini, 2019).

Figure 8.15: Most Prominent Discourse Categories in the Highly Retweeted Tweets (% of Highly Retweeted Tweets Mentioned in)

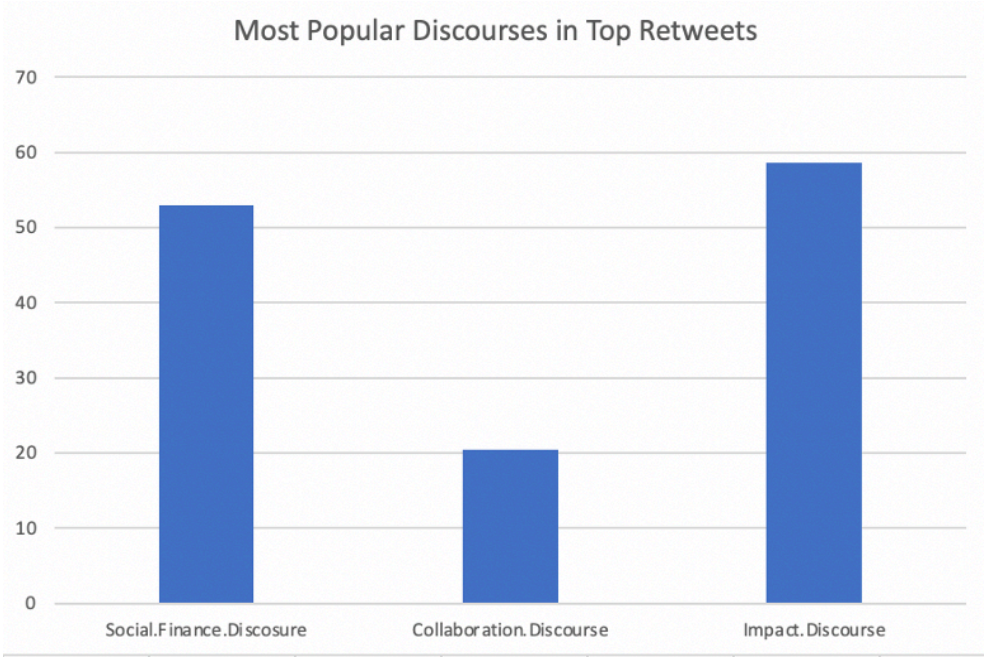


Figure 8.16: Most Prominent Sub-Categories in the Highly Retweeted Tweets (% of Highly Retweeted Tweets Mentioned in)

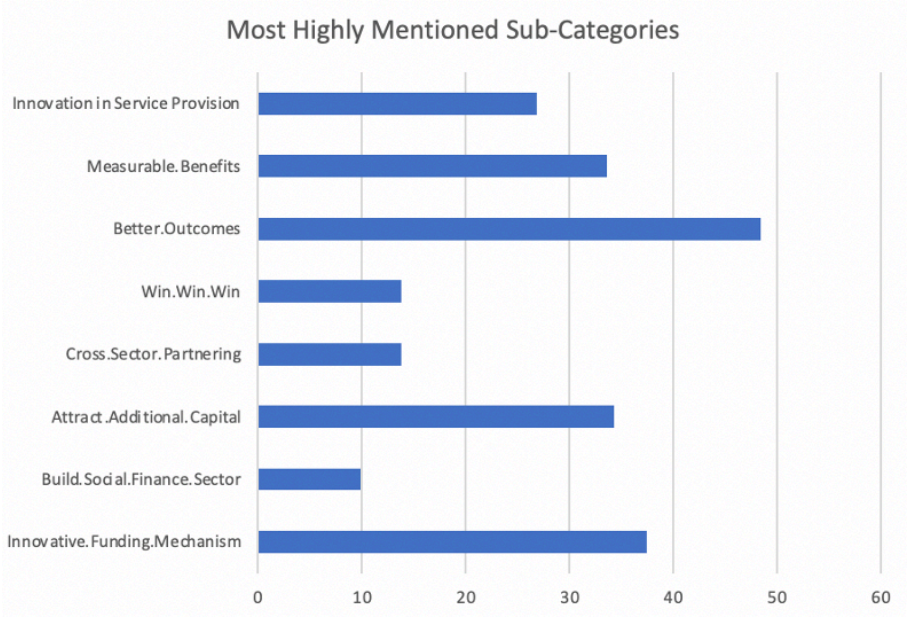


Table 8.8 provides the top 10 most highly retweeted tweets. To facilitate a comparison with other social media data, Table 8.9 includes the most highly upvoted Reddit posts (by title due to the length of the Reddit posts). The most striking observation is that the most highly retweeted tweets are overwhelmingly positive (9 out of 10), whereas the posts that have received the most 'upvotes' on Reddit are more mixed (4 positive; 1 neutral; 5 negative). This is potentially reflective of which stakeholder types are the most influential on the respective platforms. For instance, SIB stakeholders are likely to be particularly active on Twitter (especially at the organizational level). In contrast, Reddit posts are made anonymously and are less likely to be made by influential, directly involved stakeholders, such as policymakers (Amaya et al., 2019).

Table 8.8: Top 10 Most Highly Retweeted Tweets

Number of retweets	*Retweet*	*Year*	*Sentiment*
40	*"In 2019, investments from 135 #impactbonds led to jobs, better health, education, social welfare & justice in 28 countries. 👉More on @UNDP's global innovation work & how #SocialimpactBonds can support #Fin4Dev to #LeaveNoOneBehind: https://t.co/9YpjpB0M3b #SDGs"*	2019	Positive
30	*"At the beginning of the 20th century, 500,000 rhinos roamed Africa & Asia. Today, there are ~29,500 🐃 – a 41% 📉 in the past 10 yrs, due to #wildlife conservation. We're deploying #SocialImpactBonds that help conservationists to save more 🐃. #EarthDay https://t.co/hUlY8BOnUc"*	2019	Positive
23	*Tweet cannot be shared verbatim as Twitter account is no longer active. Broadly, it enthusiastically promoted the launch of the Colour Kitchen SIB, in the Netherlands.*	2015	Positive
23	*"Alla @camcomroma, presentazione dello studio dedicato ai #socialimpactbond per reinserimento persone detenute con @HUMANFdn https://t.co/sYLNtDUIp4"*	2017	Positive
21	*"The state of Connecticut launched its first #SocialImpactBond, an innovative and #sustainable financial tool. https://t.co/4MDduVbWWD"*	2017	Positive
21	*"A momentous start to 2018 as global social impact bonds surpass 100, with $392M capital raised. #108SIBsForImpact #socialimpactbonds #sibs @socialfinanceIL @SocialFinanceUS https://t.co/Mh5KeSF89M"*	2018	Positive
21	*"Pleased to share my new article on #socialimpactbonds for @apoliticalco. #impactbonds are complex — but they could help us hit the #SDGs https://t.co/cxaQqS1jJH #innovativefinance #impactinvesting @UNDP"*	2019	Positive
21	*"#ImpactBonds to reduce rhino poaching in South Africa? How are govts & @UN organizations tapping into #SocialImpactBonds to unlock #Fin4Dev & #SDGs financing? https://t.co/78sucFTTjX"*	2019	Positive
19	*"#terzosettore quando lo dicevamo noi #M5S @giosilvia86 @GiuliaDiVita @HUMANFdn @edopatriarca #socialimpactbond http://t.co/pbQDNOusxa"*	2015	Negative
19	*"More on the growing #SocialImpactBond movement here! #SIBspotlight https://t.co/pY6wl4xp0R https://t.co/QQE7xDVd4f"*	2016	Positive

Table 8.9: Top 10 Most Upvoted Reddit Posts

Number of upvotes	*Reddit post title*	*Year*	*Sentiment*
120	*"Unlike Hillary Clinton Bernie Sanders would NEVER allow Goldman Sachs to profit on "social impact bonds" based on the success of school kids"*	2016	Negative
40	*"Social Impact Bonds will create a multibillion dollar economy for startups"*	2015	Positive
40	*"TIL there's a new financial instrument called "social impact bonds" that pay you to save the gov't money by doing good deeds. Example: you save money at shelters and police stations and jails by helping the homeless get on methadone and off the street so you get paid by the bond."*	2013	Positive

32	*"Goldman Sachs' Outrageous Scheme to Profit Off Jailed Young Offenders: Too-big-to-jail Wall Street is ripping off taxpayers with a "social impact bond.""*	2014	Negative
30	*"Chicago will use close to $17 million of what are known as "social impact bonds" in addition to $4.5 million in state funds and about $10 million in capital improvement money from next year's budget to enroll more low-income children in pre-kindergarten over the next four years."*	2014	Positive
30	*"Goldman Sachs Looks to Turn a Profit on Program to Fight Recidivism: City officials say social impact bonds are a smart way to raise revenue. But government watchdogs worry about creeping corporate influence."*	2012	Neutral
29	*"Goldman Sachs' Outrageous Scheme to Profit off Jailed Young Offenders' Recidivism Rate - Too-big-to-jail Wall Street is ripping off taxpayers with a "social impact bond.""*	2014	Negative
24	*"Goldman Sachs' Outrageous Scheme to Profit Off Jailed Young Offenders: Too-big-to-jail Wall Street is ripping off taxpayers with a "social impact bond.""*	2014	Negative
21	*"CMV: The prison system should be reformed"*	2018	Positive
18	*"Social impact bonds could cost taxpayers more keep kids who need help out of care: experts"*	2019	Negative

Finally, it is also important to consider who has been sharing the posts, thus helping to spread the idea online. For the top 10 most highly retweeted tweets, Table 8.10 provides a breakdown of the percentage of retweeters with direct SIBs involvement, along with the stakeholder categories. It shows that the top retweets are largely dominated by tweets that were initially posted by influential organizations, with a relatively large social following (e.g. UNDP, BNP Paribas and Social Finance). This provides further evidence of particular organizations using their broad reach to diffuse SIBs. In some instances, the proportion of retweeters with direct SIBs experience was high, suggesting that the tweets were being shared by individuals with a vested interest in the success of SIBs. This was particularly evident in the retweet deconstructed in Table 8.11; almost all of the retweeters who shared the original post were directly involved in the Colour Kitchen SIB.

Table 8.10: Who Tweeted and Retweeted the 10 Most Highly Retweeted Tweets

#	*Year*	*# of retweets*	*Originally posted by*	*Account type of original poster*	*Is original tweeter directly involved with SIBs?*	*# SIBs directly involved with (if yes)*	*% org accounts retweeting*	*% of retweeters with direct SIBs involvement*
1	2019	40	UNDP	Organization	No	N/A	20.51%	5.13%
2	2019	30	UNDP	Organisation	No	N/A	11.54%	3.85%
3	2015	23	Colour Kitchen	Organization	Yes	1	18.18%	86.36%
4	2017	23	Andrea Orlando	Individual (politician)	No	N/A	31.82%	0%
5	2017	21	BNP Paribas CIB	Organisation	Yes	13	5.88%	17.65%
6	2018	21	Social Finance UK	Organization	Yes	25	19.04%	47.62%
7	2019	21	Gail Hurley	Individual (consultant)	Yes	Unclear	14.29%	14.29%
8	2019	21	UNDP	Organization	No	N/A	15%	0%
9	2015	19	Giulia Grillo	Individual (politician)	No	N/A	12.5%	0%
10	2016	19	SocialFinance US	Organisation	Yes	8	41.18%	47.09%

Table 8.11: Example of a Highly Retweeted Tweet – with a High Percentage of Directly Involved SIB Stakeholders who Retweeted it

Retweeter	*Type of account*	*Type of user*	*Involved with SIB/s?*	*Brief summary of involvement*
Start Foundation	Organization	Finance	Yes	Directly involved in 15 SIBs as intermediary, including the Colour Kitchen project
Erik Claus	Individual	Finance	Yes	Has worked for Stichting (involved in 3 SIBs)
Samira Zafar	Individual	Education	Yes	Has worked for Stichting (involved in 3 SIBs) - and for Society Impact, the intermediary for Colour Kitchen and a number of other SIBs
Society Impact	Organization	Intermediary/consulting	Yes	Intermediary for Colour Kitchen SIB plus 11 other SIBs
Rosanne Kamer	Individual	Public sector	Yes	Worked for Society Impact, intermediary at Colour Kitchen
RoelievanStempvoort	Individual	Finance	Yes	Works for Rabo Foundation, one of the investors in the Colour Kitchen SIB
Jorinde Hof	Individual	Public sector	Yes	Strategic account manager at the municipality of Utrecht - commissioner of the Colour Kitchen SIB
Peter Brouwer	Individual	Finance	Yes	Worked for Start Foundation, one of the investors in the Colour Kitchen SIB
SocialEnterpriseNL	Organization	NGO	No	N/A, no clear SIBs link

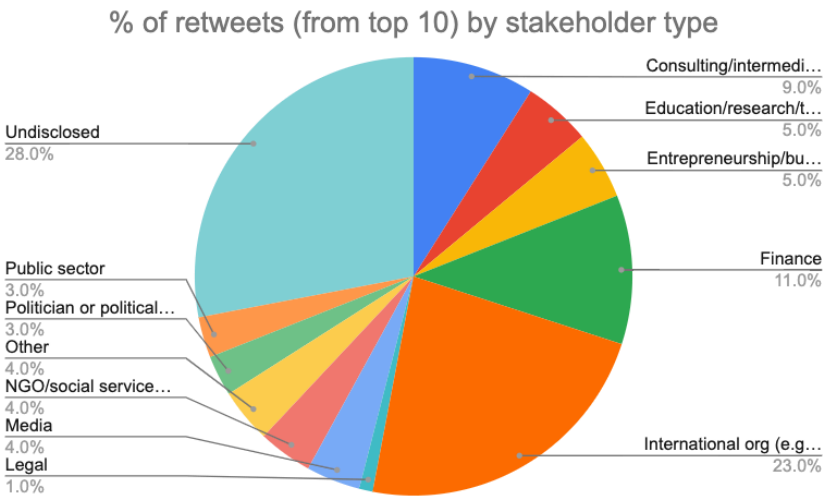
Jo de Viet	Individual	Public sector	Yes	Works for municipality of Utrecht - commissioner of the Colour Kitchen SIB
Roelant Rollingswier	Individual	Business/entrepreneurship	Yes	Worked for Stibbe (intermediary in 1 SIB project)
Martijn van Endt	Individual	Intermediary/consulting	No	N/A, no clear SIBs link
Carina ten Hoven	Individual	Finance	Yes	Project coordinator for Dutch projects at the Rabo Foundation (one of the investors in the Colour Kitchen SIB)
Lieke	Individual	Finance	Yes	Has worked for Rabobank Foundation, one of the investors in Colour Kitchen
arne flantua	Individual	Business/entrepreneurship	Yes	Co-founder of Colour Kitchen, the service provider in the Colour Kitchen SIB
Gerard Joost Esser	Individual	Undisclosed	No	N/A, no clear SIBs link
pierrevanhedel	Individual	Finance	Yes	Retired CEO Rabobank foundation, one of the investors in the Colour Kitchen SIB
Willemijn Verloop	Individual	Finance	Yes	Has been on the supervisory board for Ecorys (involved in 2 SIBs) and worked for Stichting (involved in 3 SIBs)
Rianka Habraken	Individual	Business/entrepreneurship	Yes	Worked for Rabobank, one of the investors in the Colour Kitchen SIB (and many other SIBs)
Joske Paumen	Individual	Intermediary/consulting	Yes	Worked for Colour Kitchen, the service provider in the Colour Kitchen SIB
Vorstelijk Complex	Organization	Other (arts/cultural centre)	Yes	Colour Kitchen (service provider in the Colour Kitchen SIB) is the catering partner for this organization
Jony Ferket	Individual	Political organization	Yes	Worked at Stitchting and Utrecht University

To further understand which actors have played a central role in retweeting SIBs content, I assigned each of the tweets connected to the top 10 most highly retweets to a stakeholder category. The categories that were inductively generated included:

1. Finance sector (e.g. banks, investment firms etc)
2. Consulting/intermediaries (e.g. Social Finance)
3. NGOs/social service providers/social enterprises (e.g. Colour Kitchen)
4. Public sector (e.g. civil servants, government departments etc)
5. Legal (e.g. lawyers, legal firms)
6. Entrepreneurship/business
7. International organisations (e.g. UNDP, World Bank etc)
8. Politicians or political parties
9. Media
10. Education/research/think tank
11. Other

Figure 8.17 shows the percentage of the top 10 retweets posted by each stakeholder type. While it was difficult to assign categories to many of the accounts (thus resulting in 28% being labelled as 'undisclosed'), this breakdown highlights the influential role of international organizations such as UNDP. It also demonstrates that finance stakeholders (e.g. individual investors, investing firms etc) and consultants/intermediaries have been particularly active, with each of these categories posting almost double the percentage of retweets that researchers did. Other stakeholders (e.g. public sector, politicians, service providers, media, legal) were far less represented.

Figure 8.17: Percentage of Retweets Posted by Each Stakeholder Type



8.6 Discussion

This chapter aimed to understand the persuasive techniques that SIB stakeholders have used to frame the SIB reform as 'legitimate'; and to identify the discourses most commonly used in the highly shared tweets. It has built upon prior studies in various ways. Firstly, discourse analysis has rarely been incorporated into SIB-focused studies and the only

existing study drawing on this method used press releases as the data source rather than social media data (Ormiston et al., 2020). Additionally, various types of social media data were incorporated, as a unique dataset of Reddit posts was also built, and findings were compared as a robustness check (Amaya et al., 2019). Further, while extensive literature exists focused on how reform ideas are packaged and 'sold', and how particular persuasive techniques are used, it has never been applied to the SIBs context (Abrahamson, 1991; Hood & Jackson, 1991). Existing SIB studies often assume that SIBs have been adopted for instrumental reasons (e.g. to improve efficiency, to achieve better outcomes etc) and thus do not seek to examine other possible motives which are more focused on the rhetoric and persuasive techniques used. Specifically, while Hood and Jackson (1991) provided a valuable framework for understanding how and why some ideas 'win' over others, the framework is yet to be applied to social media data. Thus, this study also makes a significant methodological contribution by demonstrating the usefulness of an existing framework (Hood & Jackson, 1991) in understanding the rise of particular public sector innovations.

Some insights from the broader literature on Twitter are integral for interpreting the findings from this chapter. For example, it was expected that the Twitter discourse (particularly in retweets) would be more positive than the broader discourse, due to factors such as positivity bias, in which positive messages tend to be shared more than negative ones, as they can *"boost the good feeling of the recipients and the public images of the sharers"* (Berger, 2014; Berger & Milkman, 2012; Kim et al., 2013; Yang et al., 2018, p. 5). Additionally, social media users tend to try to maintain positive images and 'look good' (Barasch & Berger, 2014; Yang et al., 2018). Further, novel or surprising messages are more likely to gain virality; in the context of SIBs, novelty is typically discussed or implied positively, for instance, in emphasizing that the funding model is particularly innovative

(Berger & Milkman, 2012; Kim et al., 2013). This component of the analysis does not attempt to make claims about whether the retweeting numbers are large compared to other datasets. The broader Twitter literature does not provide clear guidance on how to view retweeting numbers (e.g. what is large; relative to what etc). The number of SIBs retweets is likely much lower than those observed for less niche policy topics or issues. Instead, this chapter comments on which discourses have been more prevalent among the retweets, which ones have been overlooked and how this links to (or conflicts with) the broader SIBs and public administration literature.

The analyses presented in this chapter have addressed two specific questions. The first question asked which persuasive techniques have been used by SIB adopters to sell the idea, over time. The analysis found that most of the techniques outlined by Hood and Jackson (1991) feature highly in the Twitter dataset (Figure 8.5). Broadly involved stakeholders used the techniques most commonly, although it was hypothesized that directly involved stakeholders would deploy strategies of persuasion to the greatest extent. The growth of particular techniques over time (e.g. symmetry and time pressure) (Figure 8.7) can potentially be explained by the shift in context to developing settings and the discussion around SIBs being a funding mechanism that can assist with addressing the funding gap associated with the SDGs and the COVID-19 recovery (Alenda-Demoutiez, 2020; Mair, 2021). Many of the Twitter-specific techniques (e.g. using multiple mentions, hashtags, videos etc) were more frequently used by directly involved stakeholders (Figure 8.8), thus suggesting that they more frequently relied upon non-explicit techniques to sell the idea. Other persuasive techniques were included yet significantly under-represented in the dataset (e.g. character assassination and linking the reform to national economic performance). Importantly, this is the first known application of Hood and Jackson's (1991) 'acceptance factor' framework to a study

using social media data. Thus, it provides some methodological insights, and with further use, the framework could be refined in other studies. The framework could also be applied to a primary qualitative investigation to see if the patterns are similar in interview data. Another area for further development is the COI instrument, which was developed and applied to the random sample. Many of the actors directly involved in SIBs can benefit to a greater extent than in other policy areas (e.g. investors who can profit financially). However, there are currently no instruments specific to this context which take into account the different levels of COI among SIB stakeholders. By developing and applying this instrument, I identified that those with a medium or high COI level posted a higher percentage of positive and a lower percentage of negative tweets than those with a low or no COI. The COI instrument could be applied to different SIB studies (and more broadly, in other studies related to the diffusion of early-stage public sector innovations).

The second research question focused on identifying which discourses are prevalent in highly shared tweets; and which have been overlooked or under-represented. It was anticipated that a large proportion of the highly retweeted tweets would have been posted pre-launch and would focus on the benefits of SIB adoption. The results were consistent with this hypothesis, as a high percentage of the retweets were posted pre-implementation (e.g. announcement of SIB) and featured discourses focused on the benefits of SIB adoption, with the potential downsides often overlooked. This is aligned with the existing literature, which highlights that *"much more has been written about the promises and the processes than about the documented results of those reform efforts of the recent past"* (Hood & Dixon, 2015, p. 4). Another way in which the findings are consistent with the existing literature, is the fact that a wide array of discourses were identified (sometimes within the same tweet),

demonstrating that the idea has been packaged differently, to appeal to different stakeholders, at different times (Fraser et al., 2020; Ormiston et al., 2020).

The findings show that particular discourses (primarily focused on the promises and benefits associated with SIBs) have been heavily prioritized in the highly shared tweets. Many of the discourses identified in a previous SIBs-focused study (Ormiston et al., 2020) were common in highly retweeted tweets. However, while Ormiston et al., (2020) found that press releases primarily featured a 'social finance' and 'collaboration' discourse, the most highly shared retweets most commonly included a 'social finance' and 'impact' discourse. The 'collaboration' discourse was mentioned less frequently in tweets. Several additional categories of discourses were identified, particularly related to the possible harm and unintended consequences surrounding SIBs and the excessive hype. A key finding from the discourse analysis is the importance of identifying and analyzing discourses relating to the potential negative consequences of SIB adoption, particularly as the existence of pro-innovation biases can result in an assumption that reform adoption will inherently be beneficial. Ormiston et al., (2020) did not include any categories related to the possible downsides of SIBs. Thus, in this discourse analysis, I extended their framework to include two additional categories ('potential harm' and 'excessive hype' discourse). Other notable findings from the discursive component of this chapter included the fact that: 1) academic studies and impact evaluations rarely featured in the highly shared tweets; and 2) the 'events' mentioned primarily focus on the promises of SIBs and are focused on the 'pre' launch, rather than focusing on results.

The findings related to the question about persuasive techniques conflict slightly with the existing literature, as it was hypothesized that directly involved stakeholders (referred to

as ‘adopters’ in previous chapters) would resort to using persuasion strategies more than their counterparts who lack direct experience with SIBs. The logic of this hypothesis was that *“what makes for winning administrative doctrines is rhetorical power: the standing of the proponent and the packaging of the argument”*; those with direct experience have more to gain from the success of SIBs and thus a greater incentive to ‘package’ the idea in a particular way (Hood & Jackson, 1991, p. 11). Other scholars studying the diffusion of ideas have also noted that stakeholders who can gain from the success of the idea are likely to play a fashion setting role, by *“selecting a few administrative models from many”* and *"developing organizations' awareness and tastes for these models in order to render them fashionable and to prompt their diffusion"* (Abrahamson, 1991, p. 595). Despite this hypothesis, each persuasion strategy was identified in a greater percentage of the tweets posted by uninvolved stakeholders. This was the case for each of the possible persuasive techniques. It was only when an additional category was added (for broadly involved stakeholders such as the media) that the results appeared to align to the original hypothesis, with broadly involved stakeholders using the techniques to a greater extent than others. Thus, while the findings of this chapter have demonstrated the usefulness of the Hood and Jackson (1991) framework in explaining diffusion, they have also given some conflicting evidence on the role of particular stakeholders in using them.

The Twitter analyses provided some valuable insights as to how and why uncertainty has remained regarding the effectiveness of the SIB funding model. The lack of discussion on the role and use of evidence was particularly apparent, with SIB proponents often completely ignoring or reframing the available evidence. While tweets posted by involved stakeholders more frequently mentioned a specific evaluation, the overall percentage remained extremely low. Uninvolved stakeholders more frequently mentioned alternatives to SIBs, providing

potential evidence that SIBs have not been institutionalized, given that when institutionalized, alternatives to an idea become 'unthinkable' (Zucker, 1983, p. 5). In regards to strategies of persuasion, within the category of 'suspension of disbelief', the most common category was framing the idea as new. Concerningly, this is a strategy which assists with ensuring that existing evidence (particularly regarding negative results) is overlooked.

The above findings have important implications for three reasons. Firstly, the persuasive techniques highlighted by Hood and Jackson (1991) featured highly in the dataset (most of the techniques featured in at least a quarter of the tweets), thus suggesting that reforms are often adopted on the basis of rhetorical arguments rather than conclusive evidence. An important implication from this is that policymakers should remain aware of the persuasive techniques that are being applied, particularly in online forums and public discussions, to frame a reform idea as a valuable one to pursue. This finding links to the commentary in the existing literature on the 'malleability' of ideas, for instance, in the context of health inequalities (Smith, 2013a). Secondly, methodologically, there is a need to further test and build upon frameworks such as the one proposed by Hood and Jackson (1991). This chapter has provided some evidence of its usefulness in understanding the diffusion of public sector ideas, yet it has been underutilized in the existing literature. Refinements are required as the institutionalist field has developed considerably since it was published and the advent of social media means that there are entirely new settings and reforms to apply it to (Greenwood et al., 2017; Williams et al., 2017). Thirdly, the findings also provide further evidence of SIBs being set up with great enthusiasm yet not being adequately evaluated, with evidence rarely being discussed in online forums such as Twitter. Consequently, policymakers should avoid applying a lower threshold of rigour to 'fashionable' ideas such as SIBs, which are discussed with enthusiasm online (Abrahamson,

1991; Best, 2006). Best (2006, p. 155) emphasizes the importance of carefully considering the merits of an idea, particularly when it entails the signs of an institutional fad: *"Once we recognize institutional fads as characteristic of our culture, once we acknowledge that there already have been many such fads and that we are vulnerable to others in the future, it becomes reasonable to ask questions when a new innovation comes on the scene, to check to see where this novelty bears the hallmarks of an institutional fad".* The supply of new public policy ideas is far more limited than one might think from the 'endless parade' of *"apparently new managerial fads"*, thus underscoring the notion that it is vital to look beyond the rhetoric and online 'noise' and to more carefully consider the critiques that have been raised to date (Hood & Jackson, 1991, p. xi).

The analyses included in this chapter have several limitations. Firstly, while framework analysis was particularly relevant for this body of research, it may have led to an over-identification of the included items, as I was explicitly searching for them. Likewise, I may have overlooked items not already included in the selected frameworks. Secondly, the datasets built for and analyzed in this chapter are much smaller subsets of the entire dataset, thus making statistical analysis challenging to undertake. Nonetheless, the descriptive statistics included in this chapter (e.g. percentage of each stakeholder type, percentage of adopters etc) provided a useful starting point for analysis. Thirdly, for the analysis of the random sample, dual classification of the user type (e.g. researcher within an international organization) was not undertaken for pragmatic reasons but could be included in future empirical work. Fourth, several of the Twitter limitations highlighted in the previous chapter also apply to this one, including the problems associated with potential biases. For instance, Blank (2016, p. 679) found that *"British Twitter users are younger, wealthier, and better educated than other Internet users, who in turn are younger, wealthier, and better educated*

than the off-line British population". Consequently, any collection of tweets will be biased, and researchers have thus suggested that Twitter data are unsuitable for research in which representativeness is essential (e.g. forecasting election results) (Blank, 2016). Regardless, Twitter data can be helpful in some circumstances (relevant to this study), such as when studying elites and channels of elite influence (Blank & Sutcliffe, 2017).

While the methods applied in this chapter involved in-depth analysis of each of the included tweets (e.g. clicking through to all of the hyperlinks rather than only reading the tweets), some technical issues were associated with this approach. For instance, as the dataset consists of more than 10 years of tweets, many of the tweets contain outdated links which no longer work. Thus, in some instances, it was impossible to trace the content beyond the initial (short) tweet. While the Internet Archive was used to retrieve broken links where possible, very few of the broken links could be retrieved using this method. In most instances, the relevant website was not archived at all, or it was archived to some extent, but without the relevant page, at the appropriate time.

Additionally, a small number of stakeholders were identified who had set up blogs and regularly posted about SIBs. Two accounts tweeted about SIBs daily (e.g. to post a new blog) over a prolonged period; thus, it was evident that they would be overrepresented in the random sample. To investigate whether factors such as this had influenced the results, I conducted sensitivity analysis by removing tweets with broken links and those posted by particularly active and re-conducting the analysis (see Appendix M). As part of this, I also compared the results for English versus non-English tweets. Additionally, I compared the sentiments generated using pre-built packages (Chapter 7) to those I manually coded for the random sample, identifying some broadly similar patterns. The results of the sensitivity

analysis suggested that the findings are robust, despite the discussed methodological limitations.

8.7 Conclusion

This chapter has built upon the sentiment analysis and network analysis (Chapter 7) by using framework analysis to better understand the content of SIB-focused tweets. The analysis provides evidence of various persuasive techniques having been used by SIB stakeholders to frame the reform as a legitimate one to pursue. This was particularly evident for those who have broad rather than direct SIBs involvement (e.g. media, politicians etc). The analysis drew upon a random and purposive sample, allowing for a more in-depth qualitative analysis. This is important given that prior institutional studies have called for qualitative methods to be used to complement quantitative approaches where possible, as motivations for adoption can be challenging to identify and can change over time (Greenwood et al., 2017; Kennedy & Fiss, 2009; Lounsbury, 2007). Regarding the use of persuasive techniques, this chapter identified that the framework outlined by Hood and Jackson (1991) provides a valuable lens for analyzing how particular actors have shaped the debate. Most of the techniques (which can primarily be linked to institutional rather than instrumental drivers) featured highly in the dataset. This chapter also identified which discourses have been represented in the highly retweeted tweets. The 'social finance' and 'impact' discourses were most prevalent, with some cautionary discourses also emerging over time. Some additional insights were obtained as to how and why uncertainty has remained for so long regarding the effectiveness of the SIB funding model. For example, robust academic studies and impact evaluations were rarely mentioned in the top retweets; and the events mentioned suggest that reforms are often pursued with great enthusiasm without being adequately analyzed.

In this chapter, some particularly novel insights stem from the COI analysis, which found that stakeholders with a high COI posted far more tweets with a positive sentiment and far fewer negative tweets, on average, than those with a lesser (or no) COI. The COI analysis relates to the legitimacy-efficiency theoretical framing as I argue that in addition to varying based on timing of adoption (e.g. early adopters versus later adopters), the mix of legitimacy-efficiency drivers are influenced by the extent that a particular stakeholder has a COI. It is also related to the Hood and Jackson (1991) persuasive ideas framing, as I argue that those with a higher COI are more likely to use strategies of persuasion in their tweets, given that they have vested interests in the success of the idea. The COI analysis provides some further insight as to why evidence appears to have been overlooked in the online discourse surrounding SIBs; and why the idea is still still being promoted enthusiastically (particularly on Twitter), even when evidence has emerged of null effects and in some instances, adverse outcomes. It appears that as the level of uncertainty regarding the effectiveness of the innovation has shifted over time (i.e. more evidence has gradually emerged to counter the claims of SIB proponents), the sentiments of those with a high COI have not changed accordingly, presumably due to the vested interests.

This chapter, in isolation, provided some mixed evidence regarding instrumental and institutional drivers, and how they differ as the level of uncertainty changes. However, when combining evidence from each of the chapters (using the consilience approach), it is clear that a complex interplay of instrumental and institutional drivers has underpinned the growth of SIBs. While academic studies have increasingly critiqued the funding model (Chapter 5 and Chapter 6), these studies have rarely been discussed in the online discourse, allowing for a continued state of uncertainty, which has assisted with the continued diffusion of the idea. Ultimately, after more than 12 years of research, we see increased certainty among academics

that the idea is not fulfilling its promises. However, uncertainty remains among the wider reform community (to whom the academic certainty is not being communicated). Having presented each of the empirical chapters, I will now proceed to the concluding chapter of this thesis, which will consider how the findings from each chapter relate to each other.

Part Three: Interpretations and Contributions

Chapter 9: Discussion and concluding thoughts

9.1 Introduction

This thesis has contributed to the institutional literature on how and why organizations adopt administrative innovations (Abrahamson, 1991; Davis & Greve, 1997; Palmer et al., 1993; Westphal et al., 1997; Jonsson, 2002; Strang & Soule, 1998). Despite ongoing interest in diffusion patterns, gaps remain in understanding the interplay between economic and social considerations in public sector reform decisions, particularly when the level of uncertainty regarding intervention effectiveness changes over time. While it is now well accepted that efficiency and legitimacy drivers are not mutually exclusive, it has remained unclear how these drivers combine at different time points to propel the diffusion of early-stage public sector innovations forward. Studying how and why unproven innovations such as SIBs have rapidly (although thinly) diffused throughout various public sectors is worthy of attention because there has historically been a strong selection bias towards widely spread innovations, yet the 'in-betweener' category (i.e. neither widely spread nor un-spread) can still entail significant opportunity cost, particularly when the reform is legitimate but inefficient. This thesis has built on existing literature by incorporating novel data sources and techniques and by starting from the assumption that as the level of uncertainty surrounding intervention effectiveness changes, the mix of efficiency and legitimacy drivers also change. That is, with robust evidence and increased certainty of effectiveness (or ineffectiveness), reforms may be pursued or abandoned for instrumental reasons; conversely, poor evidence and ongoing uncertainty can assist with ideas spreading for legitimacy reasons. In the context of SIBs, I have argued that SIB proponents have benefited from the absence of evidence of efficacy and likely pursued SIBs for institutional reasons, relying on rhetoric and strategies of persuasion, rather than concrete evidence of effectiveness.

As I relied on a consilience approach and thus used different data sources and methods, this closing chapter is integral for bringing together the findings from each of the empirical chapters. In combination, the effectiveness systematic review (Chapter 5), qualitative systematic review (Chapter 6); sentiment analysis and network analysis of social media data (Chapter 7) and framework analysis of social media data (Chapter 8) provided evidence of legitimacy being central to the diffusion of SIBs, both at the start and end of the truncated diffusion curve. While Chapter 5 and Chapter 6 provided a crucial foundation, the analysis of social media data provided a particularly novel and essential contribution. The Twitter dataset is much larger than previous SIB-focused datasets and covers a more than 10-year timespan, thus allowing for consideration of how particular stakeholders have framed the idea, over time.

Unlike many other interventions, there are two broad levels of consideration when evaluating SIBs, including: 1) individual scheme level (i.e. effectiveness of individual SIB-funded interventions); and 2) delivery mechanism level (i.e. effectiveness of SIB commissioning model versus other types of commissioning) (Table 9.1). Some academics have attempted to quantitatively assess the impact of the funding model but have been unable to due to data limitations (e.g. Fraser et al., 2018). To date, existing SIB evaluations only focus on the first consideration (i.e. individual scheme level evaluations), thus preventing aggregation at the second level. With this limitation in mind and noting that SIBs have continued to spread, despite a lack of evidence, I focused on examining what explains the rapid (and wide but thin) rise of the funding model in the absence of evidence of effectiveness.

Table 9.1: SIB Evaluation - Levels of Consideration

Level	Consideration
Individual scheme level	Efficacy of individual programmes or initiatives that have been funded using a SIB structure
Delivery mechanism level	Efficacy of SIB commissioning model versus other types of commissioning

9.2 Overview of findings in relation to key research questions

In Table 9.2, I present the key findings concerning my research questions. I then discuss these findings in further detail. The empirical chapters uncovered conflicting findings regarding the possible motivations for adoption; when viewed in isolation, no single data source provides a concrete conclusion. Thus, this thesis has relied heavily on a form of 'consilience' (also known as pragmatism or pluralism) in the collection and analysis of data, as well as in the interpretation of the findings (Hood & Dixon 2010; Hood & Dixon 2015). When interpreted collectively, the findings provide evidence of legitimation being central to the diffusion of early-stage public sector innovations, such as SIBs.

Table 9.2: Summarised findings from this thesis

Chapter	Research Question	Key Findings
5	What are the effects of any intervention funded by a SIB, compared to the absence of intervention or to alternative interventions, on any social outcomes?	• No quantitative evidence available on the effect of the funding model itself. • Notable time lag in the emergence of quantitative evidence (complete lack of evidence until 2014, with non-randomized studies then slowly emerging). • Studies using randomized design and reporting positive outcomes did not emerge until 2021. • Overall, a mixed evidence base on program level outcomes; null results often reported, with few robust studies repprting positive results.
6	What are the perceptions of SIB practitioners' regarding motivations for adopting SIBs; and what does this reveal about instrumental and/or institutional drivers?	• Instrumental reasons frequently mentioned in qualitative literature but unsupported in quantitative literature. • Some legitimacy pillars more frequently represented (e.g. normative and pragmatic) than others (e.g. regulatory).
7	How have sentiments regarding SIBs changed over time on Twitter and to what extent do the	• Positive sentiments rose from 2010 – 2012; dipped between 2012 – 2015; and have increased steadily since then.

	sentiments expressed in the tweets express emerging (inconclusive) research findings?	• As negative findings have accumulated over time in the academic literature, there has not been a growth in cautionary tweets. • Nuanced (particularly negative) tweets seem to have been drowned out, across the 10+ year period.
7	Which stakeholders have played a central role in communicating and promoting the SIB idea, in online forums such as Twitter?	• Fashion-setting individuals and organizations were well-represented in the most influential accounts (assessed using centrality scores). • Many of these individuals and organizations had been directly involved in SIB contracts and thus have a vested interest in the success of the concept.
8	What discourses can be identified in the highly shared tweets; and to what extent have particular discourses dominated/others been drowned out?	• 'Social finance' and 'impact' discourses featured more highly than the 'collaboration' discourse. • Some additional (largely critical) discourses were identified in the data. • Robust academic studies and impact evaluations were rarely mentioned in the top retweets.
8	What persuasive techniques have been used by SIB stakeholders to frame the SIB funding model as a 'legitimate' reform to pursue, over time?	• Hood and Jackson's (1991) persuasive techniques were well-represented in the Twitter data. • Uninvolved stakeholders used Hood and Jackson's (1991) persuasive techniques slightly more than directly involved stakeholders, although the difference was minor. • Twitter persuasive techniques were also identified; directly involved stakeholders used most of these to a greater extent than uninvolved stakeholders. • Even more than 12 years on from the initial launch, strategies such as framing the idea as new, were frequently used.

9.2.1 What are the effects of any intervention funded by a SIB, compared to the absence of intervention or to alternative interventions, on any social outcomes?

The quantitative systematic review (Chapter 5) confirmed a severe lack of evidence regarding the effectiveness of the SIB funding model even 12 years after the introduction of the first SIB. Existing evaluations have not analyzed the impact of the funding model; some positive findings have been identified at the program level, although the findings appear to differ significantly, based on study design, with very few randomized studies showing positive results. In the qualitative systematic review, many studies referenced potential unintended consequences, for instance individuals with complex issues being ignored as they are less likely to achieve outcomes (Loxley, 2017). The chronological analysis of quantitative results demonstrated that there has been a significant time lag in results relating to intervention effectiveness (see Figures 5.4 and 5.5). Until 2014, there was a complete lack of quantitative evidence. Between 2014 – 2016, some impact evaluations were published, although they primarily showed null results and used non-randomized designs. Some positive results then emerged, but the studies still used non-randomized designs. It was not until 2021

that evaluations showed statistically significant positive results using a randomized design. However, SIBs usage had already decreased by this point, likely due to the novelty having already worn off, before positive, robust results emerged.

9.2.2 What are the perceptions of SIB practitioners' regarding motivations for adopting SIBs; and what does this reveal about instrumental and/or institutional drivers?

In Chapter 6, using a qualitative systematic review, I identified that both instrumental and institutional motives featured prominently across all time points. Some pillars of legitimacy featured more prominently in the data, particularly the normative pillar, which suggests that SIBs have often been pursued due to being perceived as the 'right thing' to do for society. While instrumental reasons for adoption were also frequently mentioned in the qualitative studies, the claims were not evidenced in the quantitative data (Chapter 5), suggesting a high level of decoupling between policy and practice. It was anticipated that a greater difference would be identified in the mix of instrumental and institutional drivers at the start versus the end of the truncated diffusion curve. However, as demonstrated in Appendix L, the top cited factors did not change drastically between the different time periods. While it seems counterintuitive for the instrumental factors to have featured as highly as they did throughout all time points and for the institutional motives to have featured from the outset (rather than only later, as predicted by Tolbert and Zucker's 2-stage diffusion model), this is consistent with the more recent institutional literature which highlights the difficulties associated with disentangling efficiency and legitimacy drivers of adoption (Kennedy & Fiss, 2009).

9.2.3 How have sentiments regarding SIBs changed over time on Twitter, and did they reflect emerging (inconclusive) research findings?

Chapter 7 used sentiment analysis of social media data to consider how sentiments regarding SIBs have changed over time. This approach which has yet to be attempted in SIB-focused studies, overcomes many of the limitations associated with methods such as citation tracking. The analysis revealed that positive tweets were far more prominent in the dataset throughout all time periods, regardless of which sentiment model was applied. This finding is consistent with many prior social media studies (Misuraca et al., 2020; Welch & Mihalcea, 2017). Tweets containing a positive sentiment were retweeted more than those with negative sentiments; and tweets posted from countries that had adopted one or more SIBs were more positive than those posted from non-adopting countries. As hypothesized, while inconclusive (and often negative) findings have accumulated in the academic literature over time, there has not been a similar growth in stakeholders tweeting negatively about the SIB funding model.

Consequently, sufficient attention has not been directed to the limitations, critiques and unintended consequences associated with SIBs. These findings, which provide strong evidence of positive sentiments being amplified and negative sentiments being overlooked in the online reform "chatter", are concerning given that many social interventions have caused significant harm despite the best intentions. Further, without critical discourse, pro-innovation biases may have been perpetuated in online forums such as Twitter (Abrahamson, 1991; Bonell et al., 2015; McCord, 2003).

9.2.4 Which stakeholders played a central role in communicating and promoting the SIB idea?

In Chapter 7, I also found strong evidence of stakeholders who stand to gain from the diffusion of SIBs having been particularly influential in spreading the idea online (Chapter 7).

Calculating various dimensions of centrality provided an alternate (previously unattempted) way to quantify the influence of different stakeholders in this field. As predicted by the management fashions literature, stakeholders from the "fashion setting network" (e.g. consultants, investors etc) ranked highly across all centrality measures (Abrahamson, 1986; DiMaggio & Powell, 1983; Kimberly, 1981; Mintzberg, 1979). In Chapter 8, I found evidence of COI level influencing sentiment; those with a medium or high level of COI (often those in the "fashion setting network") expressed a higher percentage of positive sentiments about SIBs. The COI scale should be further tested in future empirical studies; and cross-tabulated with other data sources. The Chapter 7 and Chapter 8 results support the notion that inefficient innovations can spread when fashion-setting networks promote their "faddish diffusion". I argue that this is particularly likely in the early stages of diffusion when uncertainty is inherent, yet it can continue, over time, in policy areas (such as SIBs) where high levels of complexity do not easily allow for robust evaluation.

9.2.5 What discourses can be identified in the highly shared tweets; and to what extent have particular discourses dominated/others been drowned out?

Chapter 8 analyzed the discourses contained in the most highly shared tweets. I found that the prominent discourses contained primarily focused on the promises and benefits associated with SIBs, with a much smaller proportion commenting on critiques. Rather than drawing upon concrete results, the highly shared tweets frequently mentioned events such as SIB launches rather than impact evaluations and/or academic studies. This is aligned with my original hypothesis and the existing literature, highlighting the lack of analysis focused on the results of programs as opposed to promises (Hood & Dixon, 2015, p. 4). In this chapter, I also found strong evidence of a wide array of discourses being drawn upon, often within the same tweet, thus demonstrating that the idea has been packaged differently, to appeal to different

audiences, at different time points (Fraser et al., 2020; Ormiston et al., 2020). This could be reflective of SIB proponents attempting to tailor the idea to different stakeholders, whose perceptions of what is legitimate likely differ. Additionally, I found that many of the discourses that were identified in a previous study focused on SIBs (Ormiston et al., 2020) featured highly in the dataset. However, the 'collaboration' discourse was mentioned less frequently in tweets compared to press releases (used by Ormiston et al., 2020). The 'collaboration' discourse featured highly in the qualitative systematic review (Chapter 6), suggesting that the prevalence of discourses differs slightly depending on the data source. I extended Ormiston et al's (2020) framework to include categories related to the possible downsides of SIBs, noting the importance of identifying and analyzing discourses that relate to the potential negative consequences of SIB adoption.

9.2.6 What persuasive techniques have been used by SIB adopters to frame the SIB funding model as a 'legitimate' reform to pursue, over time?

In Chapter 6, I identified that persuasive techniques feature highly in the Twitter data, thus suggesting that the strategies of persuasion outlined by Hood and Jackson (1991) provide a valuable framework for analyzing social media data. The analysis uncovered further evidence of institutional drivers having remained prevalent over time; for instance, with the idea still being framed as new even though the first SIB was launched more than 12 years ago. When breaking down the results by stakeholder type, I found that persuasive techniques have been used most commonly by broadly involved (as opposed to directly involved or uninvolved) stakeholders. This underscored the importance of considering the influence of stakeholders who may not have signed onto a SIB contract but still stand to gain from the idea's success (e.g. politicians). Directly involved stakeholders used the Twitter-specific persuasive techniques to a greater extent than those who are broadly involved or

uninvolved, thus demonstrating the importance of considering non-explicit persuasive techniques. Another key finding was that evidence has rarely been mentioned (consistent with the discourse analysis component), with mentions of SIB evaluations having decreased since 2015 and mentions of SIB alternatives having increased since 2019. I argue that the lack of a focus on outcomes and evidence across time points to the prevalence of institutional drivers, with rhetoric seemingly being more vital (and relied upon to a much greater extent) than results. Additionally, the recent increase in the mention of SIB alternatives provides further evidence that the idea has not been fully institutionalized and potentially indicates when the novelty wore off.

9.3 How have efficiency and legitimacy evolved as drivers of innovation as the level of innovation efficacy has reduced?

This thesis has sought to understand precisely how efficiency and legitimacy drivers changed as the level of uncertainty regarding innovation efficacy has reduced over time. As Chapter 5 demonstrated, SIB proponents have benefited from an extended period of uncertainty in which evidence has remained limited. I argue that this has made disillusionment and abandonment less likely than if negative findings had accumulated sooner. While some robust (yet still mixed) evidence emerged from 2019 onwards, the evidence base prior to that was characterized by studies that used weak evaluation designs and which predominantly found null or mixed results. Even now, quantitative evidence is lacking on the effect of the funding model itself (rather than only the intervention). A complete lack of evidence beyond a certain time point can suggest that an idea is not working as intended (i.e. no results, more than 12 years on is potentially indicative of low or no effectiveness). This is not a clear signal however, as the lack of results could also reflect the

fact that SIBs are difficult to evaluate and we may thus never have concrete evidence to support their efficiency or effectiveness.

Regardless, the findings have demonstrated that SIBs usage declined (from 2018 onwards) prior to the emergence of robust quantitative evidence, likely as the innovation's novelty had already worn off. The qualitative studies (Chapter 6) still frequently mentioned instrumental benefits/motivations, although this pointed towards a high level of decoupling (i.e. gap between policy and practice), thus potentially also supporting the institutional view. Unlike other reform areas, in which institutional motives have been prominent only in the later stages of diffusion (when the idea has been legitimated and is thus seen as worthy to pursue) (Tolbert & Zucker, 1983), institutional motives appear to have been particularly prominent, from the outset, in the case of SIBs.

In this thesis, I show that the lack of results, particularly over time, has created space for rhetoric (rather than results) to propel the idea forward; and that this has been prominent across all periods, for instance, with the idea continuing to be framed as new, even in the later stages of diffusion. Yet, as I have argued, legitimacy alone can only sustain an idea for so long. Thus, the emergence of more robust quantitative evidence from 2019 onwards has yet to result in a substantial upsurge in SIBs, even though the idea recently received increased political attention (Patton, 2022; Post Reporters, 2022).

9.4 What explains the truncated reform diffusion of widely but thinly diffused innovations?

In understanding how and why SIBs have spread, this thesis has sought to understand what explains the truncated reform diffusion. Empirical studies often focus on widely spread

ideas, as exemplified by the strong selection bias that exists. Yet, the "in-betweener" category of innovation (including ideas such as SIBs) can reveal valuable insights about the influence of instrumental and institutional, or efficiency and legitimacy, pressures for public management reform. The empirical chapters provided strong evidence of SIBs having a truncated reform trajectory, as the idea has often been pursued for institutional reasons. As institutional motives appear to have underpinned the growth of SIBs, there has been a particularly weak link between evidence of performance and the legitimation and diffusion of the idea.

As I identified strong evidence of SIBs continuing to spread, regardless of a lack of evidence, an important finding stemming from this research, is that uncertainty over effectiveness does not necessarily inhibit innovation when the idea carries legitimacy benefits. It is not surprising that the idea initially spread despite a lack of evidence, as innovation inherently involves some uncertainty, and the idea was likely initially protected by a 'honeymoon period' (Fichman & Levinthal, 1991). Uptake and interest in the idea dipped slightly after some negative results accumulated (e.g. failure of Riker's Island SIB), suggesting a link, to some extent, between evidence and diffusion. However, overall, the idea has received sustained attention and capital in the absence of evidence, across a more than 12-year period. Particularly in the social media data, I found that evidence has been largely overlooked, drowned out (i.e. rarely shared compared with positive tweets), and/or reframed, to maintain the perception that SIBs are a worthwhile reform to pursue. Thus, while in some instances, evidence can inform the emergence of new ideas and help policymakers to 'sell' ideas to others, in the context of SIBs, it remains underutilized (Hall, 1990, 1993; Haynes et al., 2011; Kuhn, 1962; Smith, 2013a). Regardless, legitimacy benefits may ultimately be time-limited, albeit extended by domain switching (from SIBs to DIBs).

While the weak link between evidence, legitimation and diffusion may have initially helped the SIB idea to spread absent evidence of success (as organizations may mimic others when faced with ongoing uncertainty), legitimacy alone can only propel an idea so far and for so long. At a certain point, an idea can eventually become neutral or even illegitimate due to a lack of evidence. Switching SIBs to a new field (DIBs for developing contexts) may have prolonged interest in the idea, as a new group of stakeholders became involved, and there were new people to promote the idea to. As SIBs are an example of a 'chameleonic' idea, the vagueness and elasticity of the idea has allowed it to easily shift across contexts, adapting to what different stakeholders perceive to be legitimate. The domain shift was particularly evident from the social media chapters of this thesis, which identified that many of the most highly retweeted tweets focus on DIBs rather than SIBs; and international organizations such as the UNDP have enthusiastically promoted DIBs. Without this shift, interest in the SIB reform would have likely dipped further post 2017.

The existing literature provides some indication as to why the reform has not tapered off altogether, and continues to spread, although in a truncated manner. For instance, Best (2006, p. 115) notes that later adopters are more likely to be "*interested in and motivated by knowing that many others have adopted an innovation"* rather than concrete evidence of performance. Thus, evidence may be valued less over time, and ongoing uncertainty may not necessarily be scrutinized or taken to suggest that an innovation is ineffective. Specifically, *"as an innovation spreads, the later adopters are less likely to focus on how well the novelty works, or even to understand its principles that well. In fact, in order to reach more adopters, promoters may dilute the content of their presentation to make the innovation seem easier to grasp and thus more appealing"* (Best 2006, p. 115). While some level of disillusionment was evident across the various data sources, the online rhetoric surrounding SIBs has largely

remained positive. This may not necessarily mean that SIBs are still perceived as a legitimate idea to pursue, as even when ideas are delegitimized, this can occur via a *"subtle, silent process"* (Best, 2006, p. 99). Best (2006, p. 160) notes that in comparison to 'fanfare adoption', *"disillusionment and the abandonment of novelties that aren't working well tend to be much quieter processes; nobody wants to publicize their mistakes".* Similarly, Hall (1993, p. 290) notes that *"like subatomic particles, ideas do not leave much of a trail when they shift".*

9.5 Widely but thinly diffused ideas can still play an important role even if not fully institutionalized

Regardless of whether results have been achieved and/or evidenced, the SIB funding model has played an essential role in the broader landscape, for instance, by fostering a greater interest in measuring outcomes. Even if SIBs usage continues to decline, the idea will likely persist, to some extent, potentially morphing into other funding models or approaches. From the early stages of the diffusion of SIBs, stakeholders have recognized that engaging with SIBs can produce valuable lessons for governments. For instance, Belinsky (2012) noted that *"perhaps the largest lesson, and greatest success, of this young innovation has been its ability to anchor the conversation of governments, social entrepreneurs, and impact investors around measurement, metrics, and outcomes".* Initially, it was anticipated that this broader learning and benefit (i.e. advancing the discussion on measurement) would stem from the funding model enhancing evaluation practices and encouraging more rigorous measurement. However, since then, SIBs have been strongly criticized for a lack of rigour. For instance, many academics (McHugh et al., 2013; Sinclair et al., 2014; Roy et al., 2017, p. 2) have expressed their *"surprise at the seemingly uncritical enthusiasm of both policymakers and proponents for the SIB model".* There is still a paucity of empirical evidence (Chapter 5), with many programs being framed as a success in the absence of any robust evidence; and

unsuccessful programs often receiving little attention or publicity, thus, disappearing in a *"subtle, silent"* manner (Best, 2006, p. 99). In some instances, evaluations have been announced yet never publicly released; and in others, the evaluations have had serious design flaws (as discussed in Chapter 5). Thus, ironically, while SIBs were intended to enhance evaluation practices, this has rarely occurred in practice. In many ways, it seems that they have 'anchored' the conversation around measurement by providing a model that has highlighted a stark disconnect between policy and practice in terms of robust measurement. Despite strong critiques, chapter 5 demonstrated that while existing evaluations differ considerably in terms of rigour, there are some examples of quality evaluations in this field (e.g. ID Insight, 2018). Further, while SIB evaluations have been heavily critiqued, the evaluations are often more detailed than those conducted for non-SIB-funded interventions (many of which are not evaluated at all).

9.6 Theoretical reflections

Theoretically, the findings of this thesis support the notion that multi-wave models of diffusion, centred around institutional logics, are more applicable to the context of early-stage public sector innovation. Prior institutions are important, thus making it important to consider how SIBs are linked to past reform decisions. Multi-wave models recognize that organizational practices spread, not a result of their institutionalization as individual practices, *"but rather their reliance on a shared institutional logic with practices already in place"* (Shipilov et al., 2010, p. 33). Thus, organizations may be *"investing further in the logic underlying a set of practices that they have committed to already"* (Shipilov et al., 2010, p. 33). Applied to SIBs, this suggests that the adoption of the reform was made more likely due to the prior adoption of practices sharing the same logic, particularly prior reforms aimed at 'doing more with less' (e.g. NPM) (French et al., 2022). Existing SIB studies have

discussed how SIBs reflect the logics of NPM and NPG (e.g. French et al., 2022), yet they have not explored this empirically using social media data. The discourse analysis (Chapter 8) component of this thesis found that the 'collaboration' discourse was less prevalent than the 'social finance' and 'impact' discourse; this may be because it is easier to link the SIB concept to NPM (which aligns with the 'social finance' and/or 'impact' discourse), rather than NPG (which aligns with the 'collaboration' discourse).

In addition to explaining how and why prior logics are evident in SIBs, the multi-wave approach could potentially also help to explain the subsequent growth of similar innovations, such as DIBs, in the later stages of the SIB diffusion cycle. More broadly, SIBs could also represent the latest wave of the desire to keep reforming, even in the absence of robust evidence. This is desirable as the act of reforming is often viewed as legitimacy enhancing as it shows that an organization is not complacent. For instance, Best (2006, p. 24) notes that *"we tend to become suspicious of stability, simply because stability is the absence of profess: if things aren't getting better, then there must be something wrong. No wonder institutional fads catch on – in our culture, people are eager to experience something new and better".*

Future studies could further explore the extent to which SIBs share similar logics to past reforms. Tweets regularly framed SIBs as a new concept, likely as a tactic of persuasion, thus making it difficult to identify links to prior logics in the social media data, when only looking at the explicit wording. Nonetheless, future studies using social media data could overcome this, for example, by analyzing the extent to which countries represented in the tweets had already adopted related reforms, such as NPM. Existing qualitative studies (included in the qualitative systematic review) have largely focused on current practices (and

not explored links to past logics). Thus, future qualitative interviews will be necessary to explore how multi-wave diffusion models can be applied to SIBs. For instance, semi-structured interviews could be used to explore practitioner perspectives about factors such as the extent to which SIBs are similar to prior reforms they have been involved with.

9.7 What are the implications for evidence-based policymaking?

As the SIB concept is centred around the use of evidence (as an outcomes-based funding model), it is essential to consider what the thesis findings mean in relation to EBPM. Smith (2013, p. xi) notes that in areas such as public health policy, it is *"easy for everyone to repeat the mantra of 'evidence-based policy' without being aware of the extreme complexity of many policy issues".* Likewise, in emerging policy contexts such as SIBs, where vested interests are involved, the notion of 'evidence-based' policymaking is easy (and advantageous) to promote but extremely difficult to apply in practice. This is due to various factors, such as practical evaluation challenges, a lack of incentives to properly measure ideas, and an overwhelming confidence in the effectiveness of the reform (thus resulting in an assumption that robust evaluations are not needed). More broadly, there are also limitations stemming from the difficulties associated with using an EBPM approach within the neo-liberal welfare paradigm (Roy et al., 2017). This paradigm often results in evidence being ignored and complex moral issues being reframed as technocratic issues (Roy et al., 2017). While SIB proponents often fail to properly engage with the complex ethical issues surrounding the idea, it is important to recognize that *"all public policy instruments and welfare decisions ultimately rest on some normative understanding of the 'common good'"* (Roy et al., 2017, p. 2). Consequently, society should not only be concerned about 'what works' from a narrowly economic standpoint; it is also integral to understand for whom policy instruments such as SIBs are supposed to work, and in what ways. Simply highlighting

that we do not know whether SIBs 'work' overlooks the broader research questions that must also be addressed.

Although there is currently limited evidence on the effectiveness of SIBs (as demonstrated in Chapter 5), significant resources have been devoted to attempting to understand the 'SIB effect' (FitzGerald et al., 2020). Thus, there appears to be some appetite for properly evaluating the impact of the funding model, despite the significant evaluation challenges that exist. While individual program evaluations have largely only focused on evaluating program-level outcomes (e.g. impact on recidivism), various research organizations (e.g. GO Lab) have focused their attention, across many years, on better understanding the impact of outcomes-based funding models, such as SIBs. The fact that the impact remains largely unknown highlights the very significant evaluation challenges, which exist yet are often not properly acknowledged by SIB proponents. This gap also, once again, suggests that we may never fully know if the idea works or not.

A narrow obsession with 'evidence-based policy' can be overly restrictive as, in some situations, *"research evidence must lag behind policy interventions"* (Smith, 2013, p. xi). For instance, when smoking was banned in public places in Scotland (who pioneered this public health policy), there was a lack of evidence on the efficacy of this approach. Thus, in many instances, it is more beneficial to focus on the interplay of ideas as they move between actors rather than only focusing on evidence availability. While some ideas are entirely untested when introduced, others are essentially re-circulated ideas (Smith, 2013). It is not surprising that stakeholders such as politicians frequently reframe ideas; this approach enables them to *"present an idea or policy as their own when it is in fact a re-labelling of something that has been around for quite some time"* (Smith, 2013, p. xi). SIBs have many similarities with prior

'pay for success' initiatives introduced by various public sectors. More specifically, they are a subset of PbR, a reform for which a wealth of evidence already exists (particularly concerning perverse incentives). Given that the PbR literature provides empirical evidence of perverse incentives and unintended consequences, it is problematic that SIBs have frequently been framed as a completely new idea. This has resulted in relevant evidence being overlooked and the idea being propelled forward in an environment of uncertainty (e.g. with claims that the idea is not necessarily ineffective but simply new and untested).

9.8 Limitations

This dissertation has provided a novel and timely way of better understanding the diffusion of SIBs, and of public management reform more generally; yet it has its limitations. While specific methodological weaknesses have been outlined in each of the empirical chapters, there are some overarching issues which can now be considered. Firstly, as this thesis focuses on studying an 'in-betweener' category of innovation (i.e., ideas that have not fully spread but are also not a complete 'flop'), I have primarily focused on the jurisdictions, policy areas and organizations that the SIB concept has spread to. An obvious limitation is the lack of data focused on where SIBs have *not* spread to. While the social media dataset provided a way to understand the perceptions of those who have and have not pursued SIBs, the data contained in the systematic reviews focused primarily on SIBs that had already been launched. This approach likely perpetuated the selection bias already evident in the diffusion of innovation literature, underscoring the importance of future studies focusing less exclusively on where SIBs have successfully been launched. Secondly, as I used SIBs as an example of a policy that has spread absent evidence of effectiveness, I primarily focused on the spread of the idea rather than its performance. Thus, this dissertation does not assist with understanding the 'SIB effect', a topic which remains of great interest to researchers and

policymakers alike. Finally, the sophistication of the underlying (institutional) theory presents a practical barrier as stakeholders such as policymakers may not easily relate to and use institutional research when interrogating which reform ideas to pursue (and which to overlook or abandon). As institutional theory presents a way of thinking that is unfamiliar to many practitioners, further work will be required to translate the findings from this thesis into practical tools that can inform policymaking.

9.9 Conclusion

No single data source or method provides a conclusive answer as to how efficiency and legitimacy drivers have influenced the reform process of SIBs over time, particularly as levels of uncertainty have changed. While I found some strong evidence of legitimation being central to the diffusion of SIBs, I also found evidence of other possible motives; and the interplay appears to be much more complex than envisaged in traditional models of diffusion (e.g. Tolbert & Zucker, 1983). Thus, a consilience approach was required to make sense of the often-conflicting data presented in the empirical chapters of this thesis (Hood & Dixon, 2010). In combination, these analyses show that legitimation is central to the early-stage, wide but thin, diffusion of public sector innovations; and that there are practical implications of this, such as legitimacy benefits causing decision-makers to interrogate the reform idea with lower thresholds of rigour than they otherwise would.

To build on the findings of this thesis, further empirical research is now required focused on developing and testing models of diffusion that are specifically tailored to the context of early-stage public sector innovation. These should better account for the fact that evidence may emerge post-adoption regarding ineffectiveness and/or uncertainty can remain, many years after initial adoption. The thesis findings will contribute to this future empirical

work, thus better safeguarding against the adoption of inefficient but legitimate reforms and rejection of illegitimate but efficient reforms. As I have demonstrated, while ‘in-betweener’ innovations such as SIBs may not become fully institutionalized, they can still come at a high opportunity cost and thus should be adopted with greater caution.

References

Abrahamson, E. (1986). Organizational fashions: One explanation for the evolution of technological, structural, and strategic myths. *The Academy of Management Review, 16*(3), 586-612.

Abrahamson, E. (1991). Managerial fad and fashion: the diffusion and rejection of innovations. *Academy of Management Review, 16*(3), 586-612. http://dx.doi.org/10.2307/258919.

Abrahamson, E. (1996). Management fashions. *Academy of Management Review, 21*(1), 254–285. http://dx.doi.org/10.2307/258636

Abrahamson, E. & Fairchild, G. (1999). Management fashion: lifecycles, triggers, and collective learning processes. *Administrative Science Quarterly, 44*(4), 708-740. http://dx.doi.org/10.2307/2667053

Adams, J., Hillier-Brown, F., Moore, H., Lake, A., Araujo-Soares, V., White, M., Summerbell, C. (2016). Searching and synthesising 'grey literature' and 'grey information' in public health: critical reflections on three case studies. *Systematic Reviews, 5*(1), 1-11. http://dx.doi.org/10.1186/s13643-016-0337-y

Al-Razgan, M., Alrowily, A., Al-Matham, R., Alghamdi, K., Shaabi, M., & Alssum, L. (2021). Using diffusion of innovation theory and sentiment analysis to analyze attitudes toward driving adoption by Saudi women. *Technology in Society, 65* (2021), 101558. http://dx.doi.org/10.1016/j.techsoc.2021.101558

Albertson, K., Fox, C., O'Leary, C., & Painter, G. (2018). *Payment by results and social impact bonds—Outcome-based payment systems in the UK and US*. Policy Press

Albury D. (2005). Fostering innovation in public services. *Public Money and Management*, *25*(1), 51–56. 10.1111/j.1467-9302.2005.00450.x

Aldrich, H., & Fiol, C. (1994). Fools rush in? The institutional context of industry creation. *Academy of Management Review*, 19(4), 645-670. http://dx.doi.org/10.2307/258740

Alenda-Demoutiez (2020). A fictitious commodification of local development through development impact bonds? *Journal of Urban Affairs, 42*(6), 892–906 https://doi.org/10.1080/07352166.2019.1581029

Alessandri TM and Pattit JM (2014) Drivers of R&D investment: The interaction of behavioral theory and managerial incentives. *Journal of Business Research 67*(2): 151–158

Amaya, A., Bach, R., & Kreuter, F. (2019). New Data Sources in Social Science Research: Things to Know Before Working With Reddit Data. *Social Science Computer Review, 39*(5). https://doi.org/10.1177/0894439319893

Ames, H., Glenton, C. & Lewin, S. (2019). Purposive sampling in a qualitative evidence synthesis: a worked example from a synthesis on parental perceptions of vaccination communication. *BMC Medical Research Methodology,* 19(26). https://doi.org/10.1186/s12874-019-0665-4

Anders, J., & Dorsett, R. (2017). *HMP Peterborough Social Impact Bond - cohort 2 and final cohort impact evaluation*. National Institute of Economic and Social Research. https://golab.bsg.ox.ac.uk/documents/peterborough-social-impact-bond-cohort-2-results-report.pdf

Anderson, J., Hundt, E., Dean, M., Keim-Malpass, J., & Lopez, R. (2017). "The church of online support" examining the use of blogs among family caregivers of persons with dementia. *Journal of Family Nursing*, *23*(1). https://doi.org/10.1177/1074840716681289

Anderson, P., & Tushman, M. (1991). Technological discontinuities and dominant design: A cyclical model of technological change. *Administrative Science Quarterly*, *35*(4), 604-633. http://dx.doi.org/10.2307/2393511

Anderson, R. (1999). How does the mix of routines in a population change? Technology choice in the American cement industry. In A. Miner and P. Anderson (eds.), *Advances in Strategic Management*, 16: 277-307. Greenwich, CT: JAI Press.

Ansari, S., Fiss, P., & Zajac, E. (2010). Made to fit: How practices vary as they diffuse. *Academy of Management Review*, *35*(1), 67-92. http://dx.doi.org/10.5465/AMR.2010.45577876

Apodaca, A. (1952). Corn and custom: Introduction of hybrid corn to Spanish American farmers in New Mexico. In E. H. Spicer (ed.), *Human Problems in Technological Change: A Casebook*: 35-37. New York: Russell Sage Foundation.

Arena, M., Bengo, I., Calderini, M & Chiodo, V. (2016). Social Impact Bonds: Blockbuster or Flash in a Pan? *International Journal of Public Administration, 39*(12), 927-939. http://dx.doi.org/10.1080/01900692.2015.1057852

Augier, Mie, and James G. March. (2001). Remembering Herbert A. Simon (1916–2001). *Public Administration Review, 61*(4): 396– 402. http://dx.doi.org/10.1111/0033-3352.00043

Baier, V., March, J., & Saetren, H. (1986). Implementation and ambiguity. *Scandinavian Journal of Management Studies*, *2*(3-4): 197-212. https://doi.org/10.1016/0281-7527(86)90016-2

Ball, N. (2022, August 09). Social impact bonds: what have we learned from a 10-year experiment? *Pioneers Post.* https://www.pioneerspost.com/news-views/20220809/social-impact-bonds-what-have-we-learned-10-year-experiment

Barley, S., & Kunda, G. (1992). Design and devotion: Surges of rational and normative ideologies of control in managerial discourse. *Administrative Science Quarterly*, *37*(3), 363-399.

Baum, JA., Rowley, TJ., Shipilov AV., & Chuang, Y. (2005). Dancing with strangers: Aspiration performance and the search for underwriting syndicate partners. *Administrative Science Quarterly 50*(4): 536–575. http://dx.doi.org/10.2189/asqu.50.4.536

Beck, L., Schwab, C., & Pinedo, A. Social Impact Bonds: What's in a Name? Stanford Social Innovation Review. https://ssir.org/articles/entry/social_impact_bonds_whats_in_a_name#

Béland, D. (2005) Ideas and social policy: an institutionalist perspective. *Social Policy and Administration*, *39*(1): 1–18. https://doi.org/10.1111/j.1467-9515.2005.00421.x

Belinsky, M. (2012, January 23). Social Impact Bonds: Lessons from the Field. Stanford Social Innovation Review. https://ssir.org/articles/entry/social_impact_bonds_lessons_from_the_field

Ben-Oz, C., & Greve, H. (2015). Short- and Long-Term Performance Feedback and Absorptive Capacity. *Journal of Management, 41*(7), 1827-1853. https://doi.org/10.1177/0149206312466148

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology, 24*(4), 586–607. https://doi.org/10.1016/j.jcps.2014.05.002

Berger, J., & Milkman, KL. (2012). What makes online content viral? *Journal of Marketing Research*, *49*(2), 192–205. doi:10.1509/jmr.10.0353. http://dx.doi.org/10.1509/jmr.10.0353

Best, J. (2006). *Flavor of the Month: Why Smart People Fall for Fads*. University of California Press.

Bienenstock, E., & Bonacich, P. (2021). Eigenvector centralization as a measure of structural bias in information aggregation. *The Journal of Mathematical Sociology, 46*(3), 227 - 245. http://dx.doi.org/10.1080/0022250X.2021.1878357

Biermann, J. (2016). Different meanings? The translation of inclusive education in Nigeria. *J. Inclus. Educ. Res. Dev. Educ*. *39*(2016), 10–15.

Binz, C., Harris-Lovett, S., Kiparsky, M., Sedlak, D., & Truffer, B. (2016). The thorny road to technology legitimation — Institutional work for potable water reuse in California. *Technological Forecasting & Social Change*, *103*(2016) 249–263.

Black, N. (2001). Evidence based policy: proceed with care. *BMJ: British Medical Journal*, *323*(7307), 275.

Blank, G. (2016). The Digital Divide Among Twitter Users and Its Implications for Social Research. *Social Science Computer Review*, *35*(6). https://doi.org/10.1177/0894439316671698.

Blank, G., & Sutcliffe, D. (2017, April 10). Did you consider Twitter's (lack of) representativeness before doing that predictive study? Oxford Internet Institute. https://www.oii.ox.ac.uk/news-events/news/did-you-consider-twitters-lack-of-representativeness-before-doing-that-predictive-study/

Blau, P., and Schoenherr, R. (1971). *The Structure of Organizations.* New York: Basic Books.

Bloch, C., & Bugge, M. (2013). Public sector innovation—From theory to measurement. *Structural Change and Economic Dynamics*. *27*(2013), 133-145. http://dx.doi.org/10.1016/j.strueco.2013.06.008

Blumer, H. G. (1969). Fashion: From class differentiation to collective selection. *Sociological Quarterly*, *10*(3): 275-291. http://dx.doi.org/10.1111/j.1533-8525.1969.tb01292.x

Blunkett, D. (2000). Influence or Irrelevance: can social research improve government? *Research Intelligence*, *71*(2000), 12–21.

Blyth, M. (2001). The transformation of the Swedish model: economic ideas, distributional conflict, and institutional change. *World Politics, 54*(1): 1–26.

Bolici, F., Acciarini, C., Marchegiani, L., & Pirolo, L. (2020). Innovation diffusion in tourism: how information about blockchain is exchanged and characterized on twitter. *The TQM Journal*, *2020*, 1754-2731. http://dx.doi.org/10.1108/TQM-01-2020-0016

Bonell, C., Jamal, F., Melendez-Torres, G., & Cummins, S. (2015). 'Dark logic': theorising the harmful consequences of public health interventions. *Journal of Epidemiol Community Health, 69*(1), 95–98. http://dx.doi.org/10.1136/jech-2014-204671

Borgatti, S., 2003. The key player problem. In: Breiger, R., Carley, K., * Pattison, P. (Eds.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers* (pp 241-252). Committee on Human Factors, National Research Council.

Borgatti, S., Carley, K., & Krackhardt, D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Social Networks 28*(2):124-136 DOI:10.1016/j.socnet.2005.05.001. http://dx.doi.org/10.1016/j.socnet.2005.05.001

Borgatti, Stephen P., Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. (2009). Network Analysis in the Social Sciences. *Science 323*(5916), 892–895. http://dx.doi.org/10.1126/science.1165821

Borgatti, Stephen P., Martin G. Everett, and Jeffrey C. Johnson. (2013). *Analyzing Social Networks.* Los Angeles: SAGE.

Borins, S., (2001a). Encouraging innovation in the public sector. *Journal of Intellectual Capital, 2*(3), 310–319. http://dx.doi.org/10.1108/14691930110400128

Borins, S. (2001b). Public management innovation in economically advanced and developing countries. *International Review of Administrative Sciences 67*(4), 715-31. http://dx.doi.org/10.1177/0020852301674009

Bowling, A. (2002). *Research methods in health. Investigating health and health services (2nd ed.).* Buckingham: Open University Press.

Boyd, D. and Ellison, N. (2008). Social network sites: definition, history, and scholarship. J*ournal of Computer-Mediated Communication*, *13*(1), 210-230. http://dx.doi.org/10.1111/j.1083-6101.2007.00393.x

Bozeman, B., and Kingsley, G. (1998). Risk culture in public and private organizations. *Public Administration Review, 58*(2): 109–118. http://dx.doi.org/10.2307/976358

Brennan, C., Greenhalgh, J., & Pawson, R. (2018). Guidance on guidelines: Understanding the evidence on the uptake of health care guidelines. *Journal of evaluation in clinical practice*, *24*(1), 105-116. http://dx.doi.org/10.1111/jep.12734

Brett, EI, Stevens, EM, Wagener, TL, et al. (2019), A content analysis of JUUL discussions on social media: using Reddit to understand patterns and perceptions of JUUL use. *Drug and Alcohol Dependence, 194*(2019), 358–362. http://dx.doi.org/10.1016/j.drugalcdep.2018.10.014

Britten, N., Campbell, R., Pope, C., Donovan, J., Morgan, M., & Pill, R. (2002). Using meta-ethnography to synthesise qualitative research: a worked example. *Journal of Health Services Research & Policy, 7*(4)*,* 209-215. http://dx.doi.org/10.1258/135581902320432732

Broccardo, E., Mazzuca, M., & Frigotto, M. (2019). Social impact bonds: The evolution of research and a review of the academic literature. *Corporate Social Responsibility and Environmental Management*, *27*(3), 1316-1332. https://doi.org/10.1002/csr.1886

Bromiley, P., & Washburn, M. (2011). Cost reduction vs innovative search in R&D. *Journal of Strategy and Management 4*(3), 196–214. http://dx.doi.org/10.1108/17554251111152243

Bromley, P., & Powell, W. (2012). From Smoke and Mirrors to Walking the Talk: Decoupling in the Contemporary World. *The Academy of Management Annals*, *6*(1), 483-530. http://dx.doi.org/10.5465/19416520.2012.684462

Broom, J. (2021). Social impact bonds and fast policy: Analyzing the Australian experience. *Environment and Planning A: Economy and Space, 53*(1), 113-130. https://doi.org/10.1177/0308518X2094152

Brown, S., & Eisenhardt, K. (1995). Product Development: Past Research, Present Findings, and Future Directions. *The Academy of Management Review, 20*(2), 343-378. http://dx.doi.org/10.2307/258850

Brunsson, N., & Olsen, J. (1993). *The Reforming Organization*. Volume 19. New York: Routledge.

Burches E., & Burches, M. (2020). Efficacy, Effectiveness and Efficiency in the Health Care: The Need for an Agreement to Clarify its Meaning. *International Archives of Public Health and Community Medicine, 4*(1), 035. doi.org/10.23937/2643-4512/1710035

Cabinet Office. (1999a). *Modernising Government (White Paper).* London: The Stationary Office.

Cabinet Office. (1999b). *Professional Policy Making for the Twenty-First Century*. London: Cabinet Office.

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, AL., Sandercock, P., Spiegelhalter, D., Tyrer, P (2000). Framework for design and evaluation of complex interventions to improve health. *The British Medical Journal (BMJ), 321*(7262), 694–696. http://dx.doi.org/10.1136/bmj.321.7262.694

Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, M., Donovan, J. (2003). Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. *Social Science and Medicine*, *56*(4), 671-684. http://dx.doi.org/10.1016/S0277-9536(02)00064-3

Caplen, B. (2016, January 19). Social impact investing: doing good and making money. The Banker. https://www.thebanker.com/Comment-Profiles/Editor-s-Blog/Social-impact-investing-doing-good-and-making-money

Carson, S., Pignon, C., Strickel, M., Hariharan, N., McManus, J., & Lazicky, C. (2021, March 16). *Village Enterprise Development Impact Bond Study.* AEA RCT Registry. https://doi.org/10.1257/rct.7168

Cashin-Garbutt. (2016, May 04). Finding treatments for rare diseases: an interview with Dr Richard Thompson. News Medical. https://www.news-medical.net/news/20160504/Finding-treatments-for-rare-diseases-an-interview-with-Dr-Richard-Thompson.aspx

Cavendish, C. (2019, April 05). Brexit is absorbing the oxygen needed to solve other problems. Financial Times. https://www.ft.com/content/b251eb4c-56fa-11e9-91f9-b6515a54c5b1

Chang, R. M.; Kauffman, R. J.; Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, *63*(2014), 67–80. http://dx.doi.org/10.1016/j.dss.2013.08.008

Chapman, S. (2007). *Public Health Advocacy and Tobacco Control – Making Smoking History.* Oxford: Blackwell.

Chen, D. (2012, August 02). Goldman to invest in city jail program, profiting if recidivism falls sharply. New York Times. https://www.nytimes.com/2012/08/02/nyregion/goldman-to-invest-in-new-york-city-jail-program.html

Cihon, P., & Yasseri, T. (2016). A Biased Review of Biases in Twitter Studies on Political Collective Action. *Frontiers in Physics*, *4*(34), 1-8, https://doi.org/10.3389/fphy.2016.00034.

Clark, T. (2004). The Fashion of Management Fashion: a Surge Too Far? *Organization, 11*(2):297-306. http://dx.doi.org/10.1177/1350508404030659

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, *17*(1), 1-25. http://dx.doi.org/10.2307/2392088

Compagni, A., Mele, V., & Ravasi, D. (2015). How Early Implementations Influence Later Adoptions of Innovation: Social Positioning and Skill Reproduction in the Diffusion of Robotic Surgery. *Academy of Management Journal, 58*(1)*,* 42 – 278. http://dx.doi.org/10.5465/amj.2011.1184

Conifer, D., & McKinnon, M. (2016, February 12). National Disability Insurance Scheme: Treasury considered private investors to help fund project, documents show. ABC News. https://www.abc.net.au/news/2016-02-12/treasury-considered-private-investors-to-help-fund-ndis/7161810

Conservative Party. (2010). *Big Society not Big Government.* London: Conservative Party.

Cooke A, Smith D, Booth A. (2012). Beyond PICO: the SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research, 22*(10)*,* 1435-43. doi: 10.1177/1049732312452938. http://dx.doi.org/10.1177/1049732312452938

Cooper, R. G., & Kleinschmidt, E. J. (1993) Major new products: What distinguishes the winners in the chemical industry? *Journal of Product Innovation Management, 10*(2), 90- 111. https://doi.org/10.1016/0737-6782(93)90002-8

Cooper, C., Graham, C., & Himick, D. (2016). Social impact bonds: The securitization of the homeless. *Accounting, Organizations and Society*, *55*(2016), 63-82. https://doi.org/10.1016/j.aos.2016.10.003

Cooper, C., Graham, C., O'Dwyer, B. (2013, September 22 - 25). *Social impact bonds: can private finance rescue public programs?* Accounting, Organizations and Society

Conference on “Performing Business and Social Innovation through Accounting Inscriptions”, Galway, Ireland. https://pure.strath.ac.uk/portal/en/publications/social-impact-bonds(a376dd30-45f8-4b6ba440-713a63130edc)/export.html

Craig, D., & Brooks, R. (2006). *Plundering the Public Sector: How New Labor are Letting Consultants Run off with £70 Billion of our Money*. London: Constable.

Crampton, D., Fischer, R., Richter, F., Collins, C. C., Bai, R., & Henderson, M. (2020). *Partnering for Family Success: Final Evaluation Report.* Center on Urban Poverty and Community Development. Cleveland, OH: Case Western Reserve University.

CRESSI. (2016). *CRESSI Policy Roundtable Summary: Social Innovation and Social Impact Bonds: The Role of Private Capital in the Delivery of Welfare Services*, *CRESSI Deliverable 6.3*. https://www.sbs.ox.ac.uk/sites/default/files/researchprojects/CRESSI/docs/cressi-policy-roundtable-summary-SIBs.pdf.

Critical Appraisal Skills Program. (2020). *CASP qualitative studies checklist*.https://casp-uk.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-2018_fillable_form.pdf.

Cwikel, J., Behar, L., & Rabson-Hare, J. (2000). A Comparison of a Vote Count and a Meta-Analysis Review of Intervention Research With Adult Cancer Patients. *Research on Social Work Practice, 10*(1), 139-158. http://dx.doi.org/10.1177/15527581-00010001-02

Cunningham, M., Hanson, D., Gillespie, S., Pergamit, M., Oneto, A., Spauster, P., O'Brien, T., Sweitzer, L., & Velez, C. (2021). Breaking the Homelessness-Jail Cycle with Housing First Results from the Denver Supportive Housing Social Impact Bond Initiative. Urban Institute. https://www.urban.org/sites/default/files/publication/104501/breaking-the-homelessness-jail-cycle-with-housing-first_1.pdf

Cyert, R. M., & March, J. G. (1992). *A Behavioral Theory of the Firm. 2nd ed*. Prentice Hall.

Cyert, R., & March, J. G. (1963). *A Behavioral Theory of the Firm.* New York: Wiley.

Czarniawska, B., & Joerges. B. (1996). ‘ravels of ideas. In Czarniawska, B., & Sevon, G. (Eds.), *Translating Organizational Change* (pp. 13-47). Walter de Gruyter.

Daft, R. L. (1978). A dual-core model of organizational innovation. Academy of Management Journal, 21(2), 193–210. http://dx.doi.org/10.2307/255754

Damanpour, F., Szabat, K. A. and Evan, W. M. (1989). The relationship between types of innovation and organizational performance. Journal of Management Studies, 26*(6), 587–601. http://dx.doi.org/10.1111/j.1467-6486.1989.tb00746.x*

Davies, H., & Nutley, S. (2002). *Evidence-based policy and practice: moving from rhetoric to reality (Discussion Paper 2).* St Andrews: Research Unit for Research Utilisation, University of St Andrews. https://www.researchgate.net/publication/242419804

Davis, G. F., K. A. Diekmann, and C. H. Tinsley. (1994). The decline and fall of the conglomerate firm in the 1980s: The deinstitutionalization of an organizational form. *American Sociological Review, 59*(4), 547-570. http://dx.doi.org/10.2307/2095931

Davis GF, Greve HR. (1997). Corporate elite networks and governance changes in the 1980s. American *Journal of Sociology, 103*(1), 1-37. http://dx.doi.org/10.1086/231170

De Choudhury, M, De, S (2014) Mental health discourse on Reddit: self-disclosure, social support, and anonymity. In: *Eighth international AAAI conference on weblogs and social media* (pp 71-80), AAAI Press.

de Mariz, F., & Savoia, J. (2018). Financial innovation with a social purpose: the growth of Social Impact Bonds. In: S. Boubaker, D.C umming & Nguyen, D (ed.), *Research Handbook of Investing in the Triple Bottom Line* (pp. 292-313), Edward Elgar Publishing.

Dearing, J., & Cox, J. (2018). Diffusion Of Innovations Theory, Principles, And Practice. *Health Affairs, 37(2)*, 183–190. http://dx.doi.org/10.1377/hlthaff.2017.1104

Deva, S. (1985). Effectiveness and Efficiency in Public Administration: A Theoretical Framework. *Economic and Political Weekly, 20*(35). 94-96.

Dimaggio P. Interest and agency in institutional theory. In LG. Zucker (Ed.), *Research on Institutional Patterns: Environment and Culture.* Ballinger Publishing Co. 1988

Department for Communities and Local Government (DCLG) (2015). *Qualitative Evaluation of the London Homelessness Social Impact Bond: Second Interim Report, London.* https://golab.bsg.ox.ac.uk/documents/DCLG_2015.pdf

Department for Communities and Local Government (DCLG) (2017). *The impact evaluation of the London Homelessness Social Impact Bond*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/658881/SIB_Impact_evaluation_report.pdf

Dickersin K, Scherer R, Lefebvre C. (1994). Identifying relevant studies for systematic reviews. *The British Medical Journal (BMJ), 309*(6964):1286-91. 10.1136/bmj.309.6964.1286

DiMaggio, P. & Powell, W. (1991). Introduction. In W. Powell & P. DiMaggio (Eds.) *The New Institutionalism in Organizational Analysis* (pp 1-40), University of Chicago Press.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*(2), 147-160. http://dx.doi.org/10.2307/2095101

Disley and Rubin (2014). *Phase 2 report from the payment by results Social Impact Bond pilot at HMP Peterborough, Ministry of Justice Analytical Series*. https://www.gov.uk/government/publications/phase-2-report-from-the-payment-by-resultssocial-impact-bond-pilot-at-hmp-peterborough

Disley et al (2011). *Lessons learned from the planning and early implementation of the Social Impact Bond at HMP Peterborough*. https://www.gov.uk/government/publications/social-impact-bond-at-hmp-peterborough

Disley, E., Giacomantonio, C., Kruithof, K., & Sim, M. (2015). *The payment by results Social Impact Bond pilot at HMP Peterborough*: *final process evaluation report.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/486512/social-impact-bond-pilot-peterborough-report.pdf

Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., Hsu, R., Katbama, S., Olsen, R., Smith, L., Riley, R., Sutton, AJ. (2006). Conducting a critical interpretive synthesis on the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology, 6*(35), 1-13.

Donaldson, L. (1995). *American anti-management theories of organization.* New York: Cambridge University Press.

Douglas, M. (1987). *How Institutions Think.* London: Routledge.

Downs, G. W., & Mohr, L. B. (1976). Conceptual issues in the study of innovations. Administrative Science Quarterly, *21*(4), 700-714.

Ecorys UK & ATQ Consultants. (2015). *Developing Social Impact Bonds: A Commissioner's Perspective A Case Study Report produced as part of the Commissioning Better Outcomes Evaluation.*

Ecorys UK & ATQ Consultants. (2016a). *Mental Health and Employment Partnership (MHEP) Social Impact Bond – produced as part of the CBO Fund Evaluation*. https://www.tnlcommunityfund.org.uk/media/Indepth-Reviews_MHEP_Visit-2_FINAL.pdf?mtime=20190819133237&focal=none

Ecorys UK., & ATQ Consultants. (2016b). *Reconnections Social Impact Bond: reducing loneliness in Worcestershire, Case Study Report*. https://www.bigsocietycapital.com/latest/type/research/reconnections-social-impact-bondreducing-loneliness-worcestershire

Edelman, L. B. 1992. Legal ambiguity and symbolic structures: Organizational mediation of civil rights law. *American Journal of Sociology*, 97(6), 1531–1576. http://dx.doi.org/10.1086/229939

Edmiston, D., & Nicholls, A. (2018). Social Impact Bonds: The Role of Private Capital in Outcome-Based Commissioning. *Journal of Social Policy, 47*(1), 1-20. doi:10.1017/S0047279417000125

Effective Public Health Practice Project (EPHPP). (1998). *Quality Assessment Tool For Quantitative Studies. Hamilton, ON: Effective Public Health Practice Project.* https://www.ephpp.ca/PDF/Quality%20Assessment%20Tool_2010_2.pdf

Eisner MP, Humphreys D. (2012). Measuring conflict of Interest in prevention and intervention research: A feasibility sudy. In: T. Bliesener, A. Beelmann & M. Stemmler (Eds.), *Antisocial behavior and crime: Contributions of developmental and evaluation research to prevention and intervention* (pp. 165-180). Hogrefe.

Eisner, M., Humphreys, D., Wilson, P., & Gardner, F. (2015). Disclosure of Financial Conflicts of Interests in Interventions to Improve Child Psychosocial Health: A Cross-Sectional Study. *PLoS ONE 10*(11): e0142803. http://dx.doi.org/10.1371/journal.pone.0142803

Elmore, R. F. (1991). Comment on 'Towards rigor in reviews of multivocal literatures: Applying the exploratory case study method'. Review of Educational Research, *61*(3), 293-297.

Elston, T. (2014). *Not so 'arm's length': reinterpreting agencies in UK central government*. [Doctoral dissertation, University of Nottingham].

Elston, T., & Bevan, G. (2020). Using opportunity costs to counter 'one-shot' bias in policy innovation. In: *The Palgrave Handbook of the Public Servant* (pp. 1665–1684). Palgrave Macmillan.

Enterprise Institute. (2016, March 15). *Social Impact Bonds: for (finally) virtuous collaboration between the private and public sectors.* https://medium.com/institut-de-l-entreprise/contrat-impact-social-social-impact-bond-2957e1d09640#.opjeuaft3

Erskine N, Hendricks S. (2021). The Use of Twitter by Medical Journals: Systematic Review of the Literature. *Journal of Medical Internet Research*, *23*(7):e26378. http://dx.doi.org/10.2196/26378

Etzioni, A. (1964). *Modern Organizations*. Englewood Cliffs, NJ: Prentice-Hall.

Eurobarometer Survey 359. (2011). *Attitudes on Data Protection and Electronic Identity in the European Union*. Brussels: European Commission.

Eysenbach, G. (2009): Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *Journal of Medical Internet Research*, *11(*1), e11. http://dx.doi.org/10.2196/jmir.1157

Feldman, M., & SkÖldberg, K. (2002). Stories and the rhetoric of contrariety: Subtexts of organizing (change). *Culture and Organization, 8*(4), 275–292. http://dx.doi.org/10.1080/14759550215614

Ferlie, E., J. Gabbay, L. Fitzgerald, L. Locock., & S. Dopson. (2001). Evidence-Based Medicine and Organisational Change: An Overview of Some Recent Qualitative Research. In L. Ashburner (Ed.), *Organisational Behaviour and Organisational Studies in Health Care: Reflections on the Future.* Palgrave.

Fichman, M., & Levinthal, D. A. (1991). Honeymoons and the Liability of Adolescence: A New Perspective on Duration Dependence in Social and Organizational Relationships. *The Academy of Management Review, 16*(2), 442-468. http://dx.doi.org/10.2307/258870

Fiesler, C., & Proferes, N. (2018). "Participant" perceptions of twitter research Ethics. *Social Media + Society, 4*(1), 1–14. http://dx.doi.org/10.1177/2056305118763366

Fiss, P. C., & Zajac, E. J. (2004). The diffusion of ideas over contested terrain: The (non)adoption of a shareholder value orientation among German firms. *Administrative Science Quarterly*, 49(4): 501–534. http://dx.doi.org/10.2307/4131489

FitzGerald M. (1995). The practical implications of a critique of traditional science. *International Journal of Nursing Practice, 1*(1): 2–11. http://dx.doi.org/10.1111/j.1440-172X.1995.tb00002.x

FitzGerald, C., Fraser., & Kimmitt, J. (2020). Tackling the Big Questions in Social Impact Bond Research through Comparative Approaches. *Journal of Comparative Policy Analysis: Research and Practice, 22*(2), 85-99. http://dx.doi.org/10.1080/13876988.2020.1726177

FitzGerald, C., Hameed, T., Rosenbach, F., Macdonald, J., & Dixon, R. (2021). Resilience in public service partnerships: evidence from the UK Life Chances Fund, *Public Management Review*, DOI: 10.1080/14719037.2021.2015186

Flickinger, M. (2009). The Institutionalization of Divestitures A Meta-Analysis of Stock Market Performance. Germany: Gabler.

Fombrun, C. J. (1987). Crafting an institutionally informed ecology of organizations. In G. Carroll (Ed.), *Ecological models of organizations.* Ballinger Press.

François, C. (2011). Evidential variety as a source of credibility for causal inference: beyond sharp designs and structural models. *Journal of Economic Methodology, 18*(3), 233-253, https://doi.org/10.1080/1350178X.2011.611025

Fraser, A., Tan, S., Boaz, A., & Mays, N. (2020). Backing what works? Social Impact Bonds and evidence-informed policy and practice. *Public Money & Management, 40*(3), 195-204. https://doi.org/10.1080/09540962.2020.1714303

Fraser, A., Tan, S., Kruithof, K., Sim, M., Disley, E., Giacomantonio, C., Lagarde, M., & Mays, N. (2018). *Evaluation of the Social Impact Bond Trailblazers in Health and Social Care - Final report (*PIRU Publication 2018-23). Policy Innovation Research Unit. https://golab.bsg.ox.ac.uk/documents/SIBS_Evaluation_____final_report.pdf

Fraser, A., Tan, S., Lagarde, M., and Mays, N. (2016) Narratives of Promise, Narratives of Caution: A Review of the Literature on Social Impact Bonds. *Social Policy & Administration*, *52(1),* 1-28. https://doi.org/10.1111/spol.12260

Fraser, M. W., Richman, J. M., Galinsky, M. J., & Day, S. H. (2009). *Intervention research: Developing social programs*. Oxford University Press.

Frederickson, G., Smith, K., Larimer, C., Licari, Michael. (2016). *The Public Administration Theory Primer: Third Edition.* New York, USA: Routledge Taylor & Francis.

French, M., Kimmitt, J., Wilson, R., Jamieson, D., & Lowe, T. (2022). Social impact bonds and public service reform: back to the future of New Public Management? *International Public Management Journal.* (2022)*, 1-20.* http://dx.doi.org/10.1080/10967494.2022.2050859

Friedland, R., & Alford, R. (1991). Bringing Society Back In: Symbols, Practices, and Institutional Contradictions. In W. Powell & P. DiMaggio (Eds.), *The New Institutionalism in Organizational Analysis* (pp. 232–63). University of Chicago Press.

Friedman, L. S., & Richter, E. D. (2004). Relationship between conflicts of interest and research results. *Journal of General Internal Medicine, 19(*1), 51-56. http://dx.doi.org/10.1111/j.1525-1497.2004.30617.x

Gallardo, M., Kananu, W., Lazicky, C., McManus, J., & Njogu-Ndongwe, F. (2021). *Village Enterprise Development Impact Bond Evaluation Findings.* https://www.idinsight.org/publication/village-enterprise-development-impact-bond-evaluation-findings/

Gallardo, M., Kananu, W., Lazicky, C., McManus, J., & Njogu-Ndongwe, F. (2021). *Village Enterprise Development Impact Bond Evaluation Findings.* https://www.idinsight.org/publication/village-enterprise-development-impact-bond-evaluation-findings/

Galley, A., McIsaac, E., Ymeren, J. (2014). *From Impact to Investment: The NFP Experience with Social Impact Bonds.* https://munkschool.utoronto.ca/mowatcentre/wp-content/uploads/publications/93_from_investment_to_impact.pdf

Gaylor, E., Ferguson, K., McCracken, M., Wei, X., & Spiker, D. (2016). Evaluation of child outcomes in nine Child-Parent Centers: Report for 2015-16. Prepared for IFF Pay for Success I, LLC. Menlo Park, CA: SRI International.

Gaylor, E., Ferguson, K., McCracken, M., Wei, X., & Spiker, D. (2018). Evaluation of child outcomes in nine Child-Parent Centers: Report for 2016-17. Prepared for IFF Pay for Success I, LLC. Menlo Park, CA: SRI International.

Gaylor, E., Ferguson, K., McCracken, M., Wei, X., & Spiker, D. (2019). Evaluation of child outcomes in nine Child-Parent Centers: Report for 2017-18. Prepared for IFF Pay for Success I, LLC. Menlo Park, CA: SRI International.

Gerring (2008). Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques. In J, Box-Steffensmeier., H, Brady, H., & D, Collier (Eds.), *The Oxford Handbook of Political Methodology* (pp. 645–684), Oxford University Press.

Gibbs, L., & Gambrill, E. (2002). Evidence-based practice: Counterarguments to objections. *Research on Social Work Practice, 12*(3), 452–476. http://dx.doi.org/10.1177/1049731502012003007

Gittins, R. (1988, November 02). The Madness of Strike-it-Lucky Housing Welfare. Sydney Morning Herald.

Golant, B. D. and J. A. A. Sillince (2007). The Constitution of Organizational Legitimacy: A Narrative Perspective. *Organization Studies (28)*8, 1149-1167. http://dx.doi.org/10.1177/0170840607075671

Golbeck, J. (2013). Chapter 3 - Network Structure and Measures. In: *Analyzing the Social Web* (pp. 25-44). Morgan Kaufmann.

Gondo, M. B., & Amis, J. M. (2013). Variations in practice adoption: The roles of conscious reflection and discourse. *Academy of Management Review, 38*(2), 229-247. http://dx.doi.org/10.5465/amr.2010.0312

Google Trends (2022.). "Social Impact Bonds". Retrieved May 06, 2022, from www.trends.google.com

Gough, D., Oliver, S., & Thomas, J. (Eds.). (2012). *An introduction to systematic reviews.* London: Sage Publications.

Government Outcomes Lab. (GO Lab) (2019). Educate Girls. Government Outcomes Lab. ttps://golab.bsg.ox.ac.uk/knowledge-bank/case-studies/educate-girls/

Government Outcomes Lab (GO Lab). (2023a). *Impact Bond Dataset.* Retrieved April 06, 2023 from https://golab.bsg.ox.ac.uk/knowledge-bank/indigo-data-and-visualisation/impact-bond-dataset-v2/

Government Outcomes Lab (GO Lab). (2023b). *Pipeline Dataset.* Retrieved April 06, 2023 from https://golab.bsg.ox.ac.uk/knowledge-bank/indigo/pipeline-dataset

Government Outcomes Lab (GO Lab). (2023c). *Impact Bond Network Evolution.* Retrieved April 06, 2023 from https://golab.bsg.ox.ac.uk/knowledge-bank/indigo/

Government Outcomes Lab (GO Lab). (2023d). *Impact Bond Organisation Directory.* Retrieved April 27, 2023 from https://golab.bsg.ox.ac.uk/knowledge-bank/indigo/organisation-directory/

Grady, D. (1992). Promoting Innovations in the public sector. *Public Productivity and Management Review, 16*(2): 157-71. http://dx.doi.org/10.2307/3380989

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations. *The Milbank Quarterly, 82*(4), 581–629. http://dx.doi.org/10.1111/j.0887-378X.2004.00325.x

Greenhalgh, T., & Russell, J. (2009). Evidence-Based Policymaking: a critique. *Perspectives in Biology and Medicine, 52*(2), 304–18. http://dx.doi.org/10.1353/pbm.0.0085

Greenhalgh, T. (2010). *How to read a paper: The basics of evidence-based medicine (4th ed.).* John Wiley & Sons.

Greenhalgh, T., Wherton, J., Papoutsi, C., Lynch, J., Hughes, G., A'Court, C., Hinder, S., Fahy, N., Procter, R., Shaw, S. (2017). Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. *Journal of Medical Internet Research, 19*(11):e367 http://dx.doi.org/10.2196/jmir.8775.

Greenhalgh T, Maylor H, Shaw S, Wherton J, Papoutsi C, Betton V, Nelissen N, Gremyr A, Rushforth A, Koshkouei M, Taylor J. (2020). The NASSS-CAT Tools for Understanding, Guiding, Monitoring, and Researching Technology Implementation Projects in Health and Social Care: Protocol for an Evaluation Study in Real-World Settings. *JMIR Research Protocols, 9*(5), e16861. http://dx.doi.org/10.2196/16861

Greenwood, R., Oliver, C., Suddaby, R., & Sahlin, K. (2008). *The SAGE Handbook of Organisational Institutionalism.* London, UK: SAGE Publications.

Greenwood, R., Oliver, C., Lawrence, T., & Meyer, R. (2017). *The SAGE Handbook of Organizational Institutionalism: 2nd Edition.* London, UK: SAGE Publications.

Greve, H. (1998). Performance, aspirations, and ri organizational change. *Administrative Science Quarterly, 44*: 58-86

Griffiths, R., Thomas, A., & Pemberton, A. (2016). Qualitative evaluation of the DWP Innovation Fund: Final report (Research Report No 922). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535032/rr922-qualitative-evaluation-of-the-dwp-innovation-fund-final-report.pdf

Guter-Sandu, A. (2018). *Leveraging social value: multiple valuation logics in the field of social finance.* (Unpublished Doctoral thesis, City, University of London). https://openaccess.city.ac.uk/id/eprint/20214/1/Guter-Sandu%2C%20Andrei.pdf

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., deBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., & Schünemanna. H. (2011). GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology, 64*(4), 383-394. https://doi.org/10.1016/j.jclinepi.2010.04.026

Hackman, R. (1975). Is job enrichment just a fad? Harvard Business Review, 53: 129-138.

Hall, G.E., and S.M. Hord. (1987). *Change in Schools.* Albany: State University of New York Press.

Hall, P. A. (1990) Policy paradigms, experts, and the state: the case of macroe- conomic policy-making in Britain. In S. Brooks & A.-G. Gagnon (Eds.), *Social Scientists, Policy, and the State* (pp 53-78). Praeger.

Hall, P. A. (1993). Policy paradigms, social learning, and the state: the case of economic policymaking in Britain. *Comparative Politics, 25*(3): 275–296. http://dx.doi.org/10.2307/422246

Hall, P. and Taylor, R. (1996). Political science and the three new institutionalisms. *Political Studies, 44*(5): 936–957. http://dx.doi.org/10.1111/j.1467-9248.1996.tb00343.x

Hammersley, M. (2005). Is the evidence-based practice movement doing more good than harm? Reflections on Iain Chalmers' case for research-based policy making and practice. *Evidence and Policy, 1*(1), 85–100. https://doi.org/10.1332/1744264052703203

Hannan, M., & Freeman, J. (1977). The population ecology of organizations. American Journal of Sociology, *82*(5), 340-363. http://dx.doi.org/10.1086/226424

Hannes K, Lockwood C, Pearson AA. (2010). Comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research. *Qualitative Health Research, 20*(12):1736–43. http://dx.doi.org/10.1177/1049732310378656

Hannes, K. (2011). Critical appraisal of qualitative research. In J. Noyes., A. Booth., K. Hannes., A. Harden., J. Harris., S. Lewin & C Lockwood (Eds.), Supplementary guidance for inclusion of qualitative research in Cochrane systematic reviews of interventions. Cochrane Collaboration Qualitative Methods Group.

Hannes, K., & Lockwood, C. (2012). *Synthesizing Qualitative Research*. John Wiley & Sons, Ltd.

Hansen, Jesper Asring, and Poul Aaes Nielsen. (2022). How Do Public Managers Learn from Performance Information? Experimental Evidence on Problem Focus, Innovative Search, and Change. *Public Administration Review 82*(5), 946– 57. http://dx.doi.org/10.1111/puar.13533

Hartley, J. (2005). Innovation in governance and public services: Past and present. *Public Money and Management,* 25(1), 27-34. DOI: 10.1111/j.1467-9302.2005.00447.x

Hauptman, M., Niles, J., Gudin, J., & Kaufman, H. (2021). Individual- and Community-Level Factors Associated With Detectable and Elevated Blood Lead Levels in US Children: Results From a National Clinical Laboratory. *JAMA Pediatrics. 175*(12):1252-1260. http://dx.doi.org/10.1001/jamapediatrics.2021.3518

Hawkins, R., & Bieretz, B. (2017). *Procuring for Success: Lessons to Support a Shift Toward Procuring Outcomes.* https://pfs.urban.org/library/pfs-guidance-briefs-and-reports/content/procuring-success

Hawkins, R., Mitra-Majumdar, M., McClure, D., & Paddock, E. (2017). *Using Pay for Success in Criminal Justice Projects Lessons Learned from PFS Project Stakeholders, Policy Analysis, and Landscape Analysis.* https://www.urban.org/sites/default/files/publication/94736/using-pay-for-success-in-criminal-justice-projects_0.pdf

Haynes, A. S., Derrick, G. E., Chapman, S., Redman, S., Hall, W. D., Gillespie, J. and Sturk, H. (2011). From 'our world' to the 'real world': exploring the views and behaviour of policy-influential Australian public health researchers. *Social Science & Medicine 72*(7): 1047–1055. http://dx.doi.org/10.1016/j.socscimed.2011.02.004

Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin, 8*8(2), 359–369. https://doi.org/10.1037/0033-2909.88.2.359

Hernan, MA. (2004). A definition of causal effect for epidemiological research. Journal of *Epidemiology and Community Health. 58*(4), 58: 265-271. http://dx.doi.org/10.1136/jech.2002.006361

Heusinkveld, S. (2004). *Surges and sediments.* Organization concepts between transience and continuity. Nijmegen: University of Nijmegen.

Hevenstone, D., & von Bergen, M. (2020) Public–private partnerships in Social Impact Bonds: facilitating competition or hindering transparency? *Public Money & Management, 40*(3), 205-212. http://dx.doi.org/10.1080/09540962.2020.1714304

Hevey, D. (2018). Network analysis: a brief overview and tutorial. *Health Psychology and Behavioral Medicine, 6*(1), 301-328. http://dx.doi.org/10.1080/21642850.2018.1521283

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane Handbook for Systematic Reviews of Interventions (2nd ed.).* John Wiley & Sons.

Himelboim, I., Smith, M., Rainie, L., Shneiderman, B., & Espina, C. (2017). Classifying Twitter Topic-Networks Using Social Network Analysis. *Social Media + Society, 3*(1), 1–13. http://dx.doi.org/10.1177/2056305117691545

Hirsch, P. (1997). Sociology without social structure: Neo-institutional theory meets brave new world. *American Journal of Sociology*, 102: 1702-1723.

Hood, C., & Dixon, R. (2015). *A Government that Worked Better and Cost Less? Evaluating Three Decades of Reform and Change in UK Central Government.* OUP Oxford.

Hood, C., & Jackson, M. (1991). *Administrative Argument.* Aldershot: Dartmouth Publishing.

Hood, C & and Dixon, R. (2010). The Political Payoff from Performance Target Systems: No-Brainer or No-Gainer? *Journal of Public Administration Research and Theory, 20*(2), i281–98. http://dx.doi.org/10.1093/jopart/muq022

Hopewell S, Clarke M, Lefebvre C, Scherer R. Handsearching versus electronic searching to identify reports of randomized trials. *Cochrane Database of Systematic Reviews, 2007*(2), MR000001. doi: 10.1002/14651858.MR000001.pub2. PMID: 17443625; PMCID: PMC7437388

Howison, J.; Wiggins, A.; Crowston, K. (2011). Validity Issues in the Use of Social Network Analysis with Digital Trace Data. *Journal of the Association for Information Systems 12*(12). http://dx.doi.org/10.17705/1jais.00282

Hu, Q. (2019). Twitter data in public administration: a review of recent scholarship. *International Journal of Organization Theory and Behavior, 22*(2), 209-222. http://dx.doi.org/10.1108/IJOTB-07-2018-0085

Hughes, J., Scherer, J (2014). *Foundations for Social Impact Bonds: How and Why Philanthropy Is Catalyzing the Development of a New Market.* https://thegiin.org/research/publication/foundations-for-social-impact-bonds-how-and-why-philanthropy-is-catalyzing-the

Hunter, S., Scherling, A., Cefalu, M., McBain, R. (2022). Just in Reach Pay for Success: Impact Evaluation and Cost Analysis of a Permanent Supportive Housing Program. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RRA1758-1.html.

ICF Consulting Services. (2019). *Evaluation of the Fair Chance Fund - Final Report*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793810/Fair_Chance_Fund_final_report.pdf

ID Insight (2018). *Educate Girls Development Impact Bond Final Evaluation Report*. https://golab.bsg.ox.ac.uk/documents/ID_Insight_2018_Educate_Girls_Development_Impact_Bond_-_Final_Evaluation_Report.pdf

Institute for Government. (2017, April 27). Many 'delivery units' fail to live up to expectations. https://www.instituteforgovernment.org.uk/news/latest/many-delivery-units-fail-live-expectations

Institute for Nonprofit Management. (n.d). From Innovation to Legitimacy: Innovative Activity as Legitimation Strategy in the Nonprofit Sector. WU Vienna.

https://www.wu.ac.at/en/npo/research/past-research-projeckts/from-innovation-to-legitimacy/

Iyer, DN,. & Miller, KD. (2008). Performance feedback, slack, and the timing of acquisitions. *Academy of Management Journal 51*(4): 808–822. http://dx.doi.org/10.5465/amr.2008.33666024

Jacobs, W.; Amuta, A. O.; Jeon, K. C. (2017). Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. *Cogent Social Sciences, 3*(1), 327. http://dx.doi.org/10.1080/23311886.2017.1302785

James, D., Kane, D., & Ravenscroft, C (2016). *Understanding the Capacity and need to take on Investment within the Social Sector: Case Studies.* https://www.ncvo.org.uk/images/documents/policy_and_research/funding/investment-withinthe-social-sector-2016-case-studies.pdf.

Jastania, Z., Aslam, M., Abbasi, R., & Saeedi, K. (2020). Using Social Network Analysis to Understand Public Discussions: The Case Study of #SaudiWomenCanDrive on Twitter. *International Journal of Advanced Computer Science and Applications, 11*(2). http://dx.doi.org/10.14569/IJACSA.2020.0110230

Jindal, K., & Aron, R. (2021). A systematic study of sentiment analysis for social media data. *Materials Today: Proceedings,* https://doi.org/10.1016/j.matpr.2021.01.048

Jolliffe, D., Hedderman, C. (2014). *Peterborough Social Impact Bond: Final Report on Cohort 1 Analysis*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/341684/peterborough-social-impact-bond-report.pdf

Joseph, J., Gaba, V. (2015). The fog of feedback: Ambiguity and firm responses to multiple aspiration levels. *Strategic Management Journal 36*(13): 1960–1978. https://doi.org/10.1002/smj.2333

Joy, M., & Shields, J. (2020). Debate: How do Social Impact Bonds economize social policy? *Public Money & Management, 40*(3), 190-192. http://dx.doi.org/10.1080/09540962.2020.1714300

Jung, N., & Kieser, A. (2012). Consultants in the management fashion arena. In M. Kipping & T. Clark (Eds.), *The Oxford handbook of management consulting (*pp. 327–346.), New York, NY: Oxford University Press.

Kacperczyk, A., Beckman, CM., & Moliterno, TP. (2014). Disentangling risk and change: Internal and external social comparison in the mutual fund industry. *Administrative Science Quarterly 60*(2): 228–262. http://dx.doi.org/10.1177/0001839214566297

Kaganer, E. A., S. D. Pawloski, and S. Wiley-Patton. 2(010). Building Legitimacy for IT Innovations: The Case of Computerized Physician Order Entry Systems. *Journal of the Association for Information Systems, 11*(1): 1–33. http://dx.doi.org/10.17705/1jais.00219

Katz, M. L., & Shapiro, C. 1987. R&D rivalry with licensing or imitation. *American Economic Review, 77*(3), 402– 420.

Keeling, D. (1972). *Management in Government.* London: George Allen and Unwin.

Kennedy, M., & Fiss, C. (2009). Institutionalization, Framing and Diffusion: The Logic of TQM Adoption and Implementation Decisions Among US Hospitals. *Academy of Management Journal, 52*(5), 897–91. http://dx.doi.org/10.5465/amj.2009.44633062

Kim HS, Lee S, Cappella JN, Vera L, & Emery S (2013). Content characteristics driving the diffusion of antismoking messages: Implications for cancer prevention in the emerging public communication environment. *Journal of the National Cancer Institute Monographs, 2013*(47), 182–187. http://dx.doi.org/10.1093/jncimonographs/lgt018

Kimberley, J. (1981). Managerial Innovation. In P.C. Nystrom & W.H Starbuck (Eds), *Handbook of organisational design (vol 1)* (pp. 84 – 104). New York: Oxford University Press

King, A., & Crewe, I. (2014). The Blunders of our Governments. London: Oneworld Publications.

Kingdon, J. W., & Thurber, J. A. (1984). *Agendas, alternatives, and public policies (Vol. 45).* Little, Brown Boston.

Kitto, S. C., Chesters, J., & Grbich, C. (2008). Quality in qualitative research. *Medical Journal of Australia, 188*(4), 243. http://dx.doi.org/10.5694/j.1326-5377.2008.tb01595.x

Knight, F. (1921). *Risk, uncertainty and profit.* Boston, NY: Houghton Mifflin Company.

Knight, L., & Sweeney, K. (2007). Revealing implicit understanding through enthymemes: a rhetorical method for the analysis of talk. *Medical Education, 41*(3), 226-233. https://doi.org/10.1111/j.1365-2929.2006.02681.x

Koch, P., Cunningham, P., Schwabsky, N., Hauknes, J. (2006). *Innovation in the public sector – summary and policy recommendations.* NIFU STEP, Oslo. https://nifu.brage.unit.no/nifu-xmlui/bitstream/handle/11250/226575/d24-summary-final.pdf?sequence=1

Kolleck, N. Social network analysis in innovation research: using a mixed methods approach to analyze social innovations. *European Journal of Futures Research 25*(2013). https://doi.org/10.1007/s40309-013-0025-2

Kooiman, G. (2012, August 23). Goldman invests in delinquents. Het Financieele Dagblad. https://fd.nl/ondernemen/802976/goldman-belegt-in-delinquenten

Kotiloglu, S., Chen, Y., & Lechler, T. (2019). Organizational responses to performance feedback: A meta-analytic review. *Strategic Organization, 19*(2), 1–27. http://dx.doi.org/10.1177/1476127019883361

KPMG (2014). *Evaluation of the Joint Development Phase of the NSW Social Benefit Bonds Trial*. http://www.osii.nsw.gov.au/assets/officeof-social-impact-investment/files/Evaluation-of-the-Joint-Development-Phase.pdf

Krimsky, S., Rothenberg, L. S., Stott, P., & Kyle, G. (1998). Scientific journals and their authors' financial interests: A pilot study. *Psychotherapy and Psychosomatics, 67*(4-5), 194-201. http://dx.doi.org/10.1159/000012281

Kubeczko, K., Rametsteiner, E. and Weiss, G. (2006). The role of sectoral and regional innovation systems in supporting innovations in forestry. *Forest Policy and Economics, 8*(7), 704–715. DOI:10.1016/j.forpol.2005.06.011

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Laing, K., Steer, M., Lawson, S., Penn, L., O'Brien, N., Wildman, J., & Moffatt, S. (2017). *How Ways to Wellness social prescribing is improving the health and wellbeing of people with long term conditions*. Newcastle University, Institute of Health & Society. https://golab.bsg.ox.ac.uk/documents/Laing_et_al._2017a.pdf

Lam, A. (2006). Organizational innovation. In: Fagerberg J, Mowery DC (Eds.), *The Oxford handbook of innovation* (pp. 115–147)*,* Oxford University Press.

Laudan, L. (1971). William Whewell on the consilience of inductions. *The Monist 55(*3),368–91. http://dx.doi.org/10.5840/monist197155318

Lawler, E. E. II, & Mohrman, S. A. (1985). Quality circles after the fad. *Harvard Business Review, 63*(1): 65-71.

Leahy, S., Christian, F., Ellinson, A., Eastman, C., Wakelin, D., & Darvodelsky, M. (2017). *Evaluation of the Resilient Families Service – Evaluation Stage 2 - Progress Report*

2017. https://www.osii.nsw.gov.au/assets/Uploads/_versions/1531/Info-sheet-Evaluation-of-the-Resilient-Families-Service-Progress-Report.1.pdf

Leahy, S., Christian, F., Ellinson, A., Matthew-Simmons, F., & Gomez-Bonnet, F. (2015). *Evaluation of the Resilient Families Service - Mid-term Report*. https://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/files/TBS-Evaluation-Midterm-Report.pdf

Leahy, S., Christian, F., Matthew-Simmons, F., Ellinson, A., & Gomez-Bonnet, F. (2016). *Evaluation of the Resilient Families Service - Interim Report 2016*. https://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/files/TBS-Evaluation-Interim-Report-2016.pdf

Leahy, S., Gatfield, R., Quail, S., Eastman, C., Christian, F., & Williams, I. (2020). *Evaluation of the Resilient Families Service - Final Report 2020*. https://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/Evaluation-of-the-Resilient-Families-Service-Final-Evaluation-April-2020.pdf

Lehman, DW., & Hahn, J. (2013). Momentum and organizational risk taking: Evidence from the National Football League. *Management Science 59*(4): 852–868. http://dx.doi.org/10.1287/mnsc.1120.1574

Lin, W-T. (2014). How do managers decide on internationalization processes? The role of organizational slack and performance feedback. *Journal of World Business 49*(3): 396–408. http://dx.doi.org/10.1016/j.jwb.2013.08.001

Lewin, S., Glenton, C., Munthe-Kaas, H., Carlsen, B., Colvin, CJ., Gülmezoglu, M., Noyes, J., Booth, A., Garside, R., Rashidian, A. (2015). Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). *PLOS Medicine, 12*(10), e1001895. http://dx.doi.org/10.1371/journal.pmed.1001895

Liberati, A., Altman, DG., Tetzlaff, J., Mulrow, C., Gøtzsche, P., Ioannidis, J., Clarke, M., Devereaux., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ (Clinical research ed), 339*(jul21 1), b2700. http://dx.doi.org/10.1136/bmj.b2700

Liebman, J. (2013). Impact. *Brookings Institute; Building on Recent Advances in Evidence-Based Policymaking.* Brookings Institute. http://www.brookings.edu/research/building-on-recent-advances-inevidence-based-policymaking

Lindblom, C. E. (1959). The science of" muddling through. *Public Administration Review, 19*(2), 79-88. http://dx.doi.org/10.2307/973677

Lindblom, C.E. (1968) *The policy-making process.* Englewood Cliffs, NJ: Prentice Hall.

Lindblom, C. E., & Woodhouse, E. J. (1980). *The policy-making process.* Pearson.

Littell JH, Corcoran JC, Pillai V (2008). *Systematic Reviews and Meta-Analysis.* Oxford University Press.

Lizarondo, L., Stern, C., Apostolo, J., Carrier, J., Borges, K., Godfrey, C., Kirkpatrick, P., Pollock, D., Rieger, K., Salmond, S., Vandyk, A., & Loveday, H. (2022). Five common pitfalls in mixed methods systematic reviews: lessons learned. *Journal of Clinical Epidemiology*, 148, 178. http://dx.doi.org/10.1016/j.jclinepi.2022.03.014

Lockwood C, Munn Z, Porritt K. (2015). Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. *International Journal of Evidence-Based Healthcare, 13*(3):179–87. http://dx.doi.org/10.1097/XEB.0000000000000062

Locock, L., S. Dopson, D. Chambers, and J. Gabbay. (2001). Understanding the Role of Opinion Leaders in Improving Clinical Effectiveness. *Social Science & Medicine 53*(6), 745–57. http://dx.doi.org/10.1016/S0277-9536(00)00387-7

Lomax, A. (2015). Social Impact Bonds: An Exciting New Way to Invest. *Fox Business.* https://www.foxbusiness.com/markets/social-impact-bonds-an-exciting-new-way-to-invest

Lorenc, T., Petticrew, M., Whitehead, M., Neary, D., Clayton, S., Wright, K., Thomson, H., Cummins, S., Sowden, A., & Renton, A. (2014). Crime, fear of crime and mental health: synthesis of theory and systematic reviews of interventions and qualitative evidence. Public Health Research, 2(2), 1-398. http://dx.doi.org/10.3310/phr02020

Loughran, T., & McDonald, B. (2010). When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), (35-65). http://dx.doi.org/10.1111/j.1540-6261.2010.01625.x

Lounsbury, M. (2007). A tale of two cities: Competing logics and practice variation in the professionalizing of mutual funds. *Academy of Management Journal, 5*0(2), 289-330. http://dx.doi.org/10.5465/amj.2007.24634436

Lounsbury, M. (2002). Institutional transformation and status mobility: The professionalization of the field of finance. *Academy of Management Journal, 45*(1), 255–266. http://dx.doi.org/10.2307/3069295

Lounsbury, M., & Beckman, CM. (2015). Celebrating organization theory. *Journal of Management Studies 5*2(2): 288–308. http://dx.doi.org/10.1111/joms.12091

Loveless, L. (2020). *Evaluation of the Life Chances Fund - Interim report.* ICF Consulting Services. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/960511/ICF_interim_report_final_-_V2.pdf

Loxley, J. (2017). *Social Impact Bonds and the Financing of Child Welfare.* https://policyalternatives.ca/sites/default/files/uploads/publications/Manitoba%20Office/2017/07/SIBs%20and%20Child%20Welfare.pdf

Loxley, J., & Hajer, J. (2019). Public–private partnerships, social impact bonds, and the erosion of the state in Canada. *Studies in Political Economy, 100*(1), 18-40. http://dx.doi.org/10.1080/07078552.2019.1612167

Lucas, GJ., Knoben, J., & Meeus, MT. (2018) Contradictory yet coherent? Inconsistency in performance feedback and R&D investment change. *Journal of Management 44*(2), 658–681. https://doi.org/10.1177/0149206315584821

MacRoberts, M., & MacRoberts, B. (1996). Problems of citation analysis. *Scientometrics, 36*(3), 435–444. http://dx.doi.org/10.1007/BF02129604

MacVaugh, J., & Schiavone, F. (2010). Limits to the diffusion of innovation: A literature review and integrative model. *European Journal of Innovation Management, 13*(2):197-221, http://dx.doi.org/10.1108/14601061011040258

Maddock, S. (2008). *Creating the conditions for public service innovation*, National School of Government, London. https://www.researchgate.net/publication/264541547_Creating_the_conditions_for_public_innovation

Madsen, D., & Slåtten, K. (2015). The balanced scorecard: Fashion or virus? *Administrative Sciences, 5*(2), 90–124. http://dx.doi.org/10.3390/admsci5020090

Mahood, Q., Van Eerd, D., Irvin, E. (2014). Searching for grey literature for systematic reviews: challenges and benefits. *Research Synthesis Methods, 5*(3):221–34. http://dx.doi.org/10.1002/jrsm.1106

Mair, V. (2021, June 24). Could social impact bonds make a comeback as part of the COVID-19 recovery? Responsible investor. https://www.responsible-

investor.com/could-social-impact-bonds-make-a-comeback-as-part-of-the-covid-recovery/

March, J. and Olsen, J, eds. (1976). *Ambiguity and Choice in Organisations.* Bergen, Norway: Universitetsforlaget.

Mason, P., Lloyd, R., Andrews, M., & Henry, N. (2014). *Learning from the Development and Commissioning of the London Homelessness SIB: A Summary Report from the Qualitative Evaluation.* Department for Communities and Local Government. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/658939/Commissioning_Social_Impact_Bonds.pdf

Maze. (2018). The Learnings and Outcomes of the Pilot Project "Social Impact Bond Junior Code Academy " Executive Summary April 2018. https://maze-impact.com/wp-content/uploads/2018/05/RESULTADOS_TIS_EN_v2.pdf

McConnell, M.A., Zhou, R.A., Martin, M.W., Gourevitch, R.A., Steenland, M., Bates, M.A,, Zera, C., Hacker, M., Chien, A., Baicker, K. (2020). Protocol for a randomized controlled trial evaluating the impact of the Nurse-Family Partnership's home visiting program in South Carolina on maternal and child health outcomes. *Trials, 21*(1), 997. https://doi.org/10.1186/s13063-020-04916-9

McConnell, M., Rokicki, S., Ayers, S., Allouch, F., Perreault, N., Gourevitch, R., Martin, M., Zhou, A., Zera, C., Hacker, M., Chien, A., Bates, M., & Baicker, K. (2022). Effect of an Intensive Nurse Home Visiting Program on Adverse Birth Outcomes in a Medicaid-Eligible Population. *JAMA, 328*(1). http://dx.doi.org/10.1001/jama.2022.9703. AMA,.

McCord. (2003). Cures That Harm: Unanticipated Outcomes of Crime Prevention Programs. *The Annals of the American Academy of Political and Social Science, 587*(1), 16-30. http://dx.doi.org/10.1177/0002716202250781

McHugh, N., Sinclair, S., Roy, M. J., Huckfield, L., & Donaldson, C. (2013). Social Impact Bonds: A Wolf in Sheep's Clothing? *Journal of Poverty and Social Justice, 21*(3), 247–257. http://dx.doi.org/10.1332/204674313X13812372137921

McLennan, G. (2004) Travelling with vehicular ideas: the case of the third way. *Economy and Society 33*(4), 484–499. http://dx.doi.org/10.1080/0308514042000285251

McNeil, K., Brna, P. M. and Gordon, K. E. (2012). Epilepsy in the Twitter era: a need to retweet the way we think about seizures. *Epilepsy & Behavior, 23*(2), 127- 130. http://dx.doi.org/10.1016/j.yebeh.2011.10.020

Medaglia, R. and Zhang, L. (2017), Mapping government social media research and moving it forward: a framework and a research agenda. *Government Information Quarterly, 34*(3), 496-510. http://dx.doi.org/10.1016/j.giq.2017.06.001

Medical Research Council. (2000). *MRC's Framework for the Development and Evaluation of RCT's for Complex Interventions.* London: Medical Research Council.

Medvedev, AN, Lambiotte, R, Delvenne, JC (2019) The anatomy of Reddit: an overview of academic research. In: Ghanbarnejad, F, Saha Roy, R, Karimi, F, et al. (Eds.), *DOOCN 2017: Dynamics On and of Complex Networks III* (183-204), Springer.

Mehta, J. (2011) The varied roles of ideas in politics: from 'whether' to 'how'. In D. Béland & R. H. Cox (Eds.), *Ideas and Politics in Social Science Research* (23-46). Oxford University Press.

Melnyk, B. M., & Fineout-Overholt, E. (2015). *Evidence-based practice in nursing & healthcare: a guide to best practice.* Philadelphia, PA: Wolters Kluwer Health.

Mergel, K., Rethemeyer, K., & Isett, K., (2016). Big Data in Public Affairs. *Public Administration Review, 76(*6), 928–937. http://dx.doi.org/10.1111/puar.12625

Metaxas, P., Mustafaraj, E., Wong, K., Zeng, L., O'Keefe, M. and Finn, S. (2015). What do retweets indicate? Results from user survey and meta-review of research. In: *Ninth International AAAI Conference on Web and Social Media* (658-661). IEEE Press.

Meyer, J.W., & Rowan, B. (1977/1991). Institutionalised organisations: formal structure as myth and ceremony. *American Journal of Sociology, 83*(2), 440-463.

Mhaskar, R., Emmanuel, P., Mishra, S., Patel, S., Naik, E., & Kumar, A. (2009). Critical appraisal skills are essential to informed decision-making. *Indian Journal of Sexually Transmitted Diseases, 30*(2), 112–119. doi: 10.4103/2589-0557.62770

Ministry of Justice (2014). *Peterborough Social Impact Bond HMP Doncaster - Payment by Results pilots - Final reconviction results for cohorts 1.* https://www.bl.uk/collection-items/peterborough-social-impact-bond-hmp-doncaster-payment-by-results-pilots-final-reconviction-results-for-cohorts-1

Mintrom, M., and Vergari, S. (1998). Policy networks and innovation diffusion: the case of state education reforms. *The Journal of Politics, 60*(1), 126–148. http://dx.doi.org/10.2307/2648004

Mintzberg, H. (1979). *The structuring of organizations.* New York: Prentice-Hall.

Misuraca, M., Forciniti, A., Scepi, G., & Spano, M. (2020). Sentiment Analysis for Education with R: packages, methods and practical applications. *Software Review.* https://doi.org/10.48550/arXiv.2005.12840

Moffatt, S., Steer, M., Lawson, S., Penn, L., & O'Brien, N. (2017). Link Worker social prescribing to improve health and well-being for people with long-term conditions: qualitative study of service user perceptions. *BMJ Open, 7*(7), e015203. http://dx.doi.org/10.1136/bmjopen-2016-015203

Moher D., Liberati A., Tetzlaff J., Altman DG & The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine, 6*(7), e1000097. http://dx.doi.org/10.1371/journal.pmed.1000097

Moran, T. (2010). *Empowering change: Fostering innovation in the Australian public sector.* Attorney-General's Department. https://catalogue.nla.gov.au/Record/4901069

Morrell, K., & Hewison, A. (2013). Rhetoric in policy texts: the role of enythmeme in Darzi's review of the NHS. *Policy & Politics, 4*1(1), 59-79. http://dx.doi.org/10.1332/030557312X645801

Moser, A., Puhan, M., & Zwahlen, M. (2020). The role of causal inference in health services research II: a framework for causal inference. *International Journal of Public Health 65*(2020), 367–370. https://doi.org/10.1007/s00038-020-01334-1

Mulgan, G. (2007). *Ready or not? Taking innovation in the public sector seriously.* NESTA. www.nesta.org.uk/ready-or-nottaking-innovation-in-the-public-sector-seriously

Mulgan, G., & Albury, D. (2003). *Innovation in the Public Sector.* Strategy Unit, Cabinet Office. http://www.sba.oakland.edu/faculty/mathieson/mis524/resources/readings/innovation/innovation_in_the_public_sector.pdf

Munn Z, Tufanaru C, Aromataris EJBI. (2014). Systematic reviews: data extraction and synthesis. *AJN, American Journal of Nursing, 114*(7), 49–54. http://dx.doi.org/10.1097/01.NAJ.0000451683.66447.89

Munn, Z., Stern, C., Aromataris, E., Lockwood, A., & Jordan, Z. (2018). What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Medical Research Methodology, 18*(1). https://doi.org/10.1186/s12874-017-0468-4

National Audit Office (NAO). (2012). *Efficiency and reform in government corporate functions through shared service centres.* https://www.nao.org.uk/wp-content/uploads/2012/03/10121790es.pdf

Naveh, E., Meilich, O., & Marcus, A. (2006). The effects of administrative innovation implementation on performance: an organizational learning approach. *Strategic Organization, 4*(3), 275–302. http://dx.doi.org/10.1177/1476127006066969

Neiger, B. L., Thackeray, R., Burton, S. H., Giraud-Carrier, C. G., & Fagen, M. C. (2013). Evaluating social media's capacity to develop engaged audiences in health promotion settings: use of Twitter metrics as a case study. *Health Promotion Practice, 14*(2), 157-162. http://dx.doi.org/10.1177/1524839912469378

Neilson, S. (2001). *IDRC supported research and its influence on public policy: Knowledge utilization and public policy processes: A literature review.* Ottawa: International Development Research Centre. https://www.researchgate.net/publication/277188298_IDRC-supported_research_and_its_influence_on_public_policy_knowledge_utilization_and_public_policy_processes_a_literature_review/citation/download

Neyland, D. (2017). On the transformation of children at-risk into an investment proposition: A study of Social Impact Bonds as an anti-market device. *The Sociological Review 2018, 66*(3) 492–510. http://dx.doi.org/10.1177/0038026117744415

Ngai, E.W., Moon, K.-l.K., Lam, S.S., Chin, E.S. & Tao, S.S. (2015). Social media models, technologies, and applications: an academic review and case study. *Industrial Management & Data Systems, 115*(5). 769-802. http://dx.doi.org/10.1108/IMDS-03-2015-0075

Noyes, J., & Lewin, S. (2011). Chapter 5: Extracting Qualitative Evidence. In J. Noyes, A. Booth, K. Hannes, A. Harden, J. Harris, S. Lewin, & C. Lockwood (Eds.), *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions: Cochrane Collaboration Qualitative Methods Group*.

NSW Treasury. (2015). Social Impact Investment Policy. Office of Social Impact Investment. https://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/files/Social-Impact-Investment-Policy.pdf

NSW Treasury. (2022). Social Impact Investment Policy 2.0: Growing our impact. Office of Social Impact Investment. https://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/files/Final-NSW-Social-Impact-Investment-Policy-2.0.pdf

Nutley, S. M., Smith, P. C., & Davies, H. T. (2000). *What works?: Evidence-based policy and practice in public services.* Bristol: Policy Press.

Nutley, S. M., Walter, I., & Davies, H. T. (2007). *Using evidence: How research can inform public services.* Bristol: Policy Press.

Nystrom, P. C., & Starbuck, W. H. (1984). Organizational facades. *Academy of Management Proceedings, 984*(1),182-185. http://dx.doi.org/10.5465/ambpp.1984.4978761

Oakley, A. (1998). Experimentation and social interventions: a forgotten but important history. *BMJ, 317*(7167), 1239–1242. http://dx.doi.org/10.1136/bmj.317.7167.1239

Oakley, A. (2002). Social Science and Evidence-based Everything: the case of education. *Educational Review, 54*(3). http://doi.org/10.1080/0013191022000016329

Oakley, A., Gough, D., Oliver, S., & Thomas, J. (2005). The politics of evidence and methodology: lessons from the EPPI-Centre. *Evidence & Policy: A Journal of Research, Debate and Practice, 1*(1), 5-32. http://dx.doi.org/10.1332/1744264052703168

Ogilvie, D., Egan, M., Hamilton, V., & Petticrew, M. (2005). Systematic reviews of health effects of social interventions: 2. Best available evidence: how low should you go?

Journal of Epidemiology & Community Health, 59(10), 886–892. http://dx.doi.org/10.1136/jech.2005.034199

O'Kelly, M. (2014, August 08). Encouraging fall in reconvictions. Financial Times. https://www.ft.com/content/770aff00-1d8f-11e4-8f0c-00144feabdc0

Oliver, C. (1991). Strategic responses to institutional processes. *Academy of Management Review, 16*(1), 145-179. http://dx.doi.org/10.2307/258610

Ooi, E. J., Poynton, S., & Halloran, N. (2020). *Social impact investment and recidivism: A field experiment with high-risk parolees (Crime and Justice Bulletin No. 234).* NSW Bureau of Crime Statistics and Research. https://www.bocsar.nsw.gov.au/Pages/bocsar_publication/Pub_Summary/CJB/cjb234-Social-impact-investment-and-recidivism.aspx

Ormiston, J., Charlton, K., Donald, S., & Seymour, R. (2015). Overcoming the Challenges of Impact Investing: Insights from Leading Investors. *Journal of Social Entrepreneurship, 6*(3), 352-378. http://dx.doi.org/10.1080/19420676.2015.1049285

Ormiston, J., Moran, M., Castellas, E., & Tomkinson, E. (2020). Everybody wins? A discourse analysis of competing stakeholder expectations in Social Impact Bonds. *Public Money & Management, 40*(3), 237-246. http://dx.doi.org/10.1080/09540962.2020.1714316

Oroxom, R., Glassman, A., & McDonald, L. (2018). *Structuring and Funding Development Impact Bonds for Health: Nine Lessons from Cameroon and Beyond. Center for Global Development.* https://www.cgdev.org/publication/structuring-funding-development-impact-bonds-for-health-nine-lessons

Osborne, Stephen P. (1998). *Voluntary organizations and innovation in public services.* London: Routledge.

Pagoto, S., Nebeker, C. (2019). How scientists can take the lead in establishing ethical practices for social media research. *Journal of the American Medical Informatics Association 26*(4): 311–313. http://dx.doi.org/10.1093/jamia/ocy174

Painter, G., Albertson, K., Fox, C., & O'Leary, C. (2018). *Social Impact Bonds: More Than One Approach.* Stanford Social Innovation Review. https://ssir.org/articles/entry/social_impact_bonds_more_than_one_approach#

Palmer, Donald, P. Devereaux Jennings, and Xueguang Zhou. (1993). Late adoption of the multidivisional form. *Administrative Science Quarterly*, *38*(1993), 100–31.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval, 2*(1-2), 1–135. http://dx.doi.org/10.1561/1500000011

Parker, ON., Krause, R., Covin, JG. (2017). Ready, set, slow: How aspiration-relative product quality impacts the rate of new product introduction. *Journal of Management 43*(7), 2333–2356. http://dx.doi.org/10.1177/0149206315569314

Parsons, J., Weiss, C., Wei, Q. (2016). *Impact Evaluation of the Adolescent Behavioral Learning Experience (ABLE) Program.* Vera Institute of Justice. https://www.vera.org/downloads/publications/rikers-adolescent-behavioral-learning-experience-evaluation.pdf

Parvensky, J., & Clark, C. (2018, November 28). The Podium: 'Caring 4 Denver' gives the homeless a new lease on life. *Colorado Politics.* https://www.coloradopolitics.com/opinion/the-podium-caring-4-denver-gives-the-homeless-a-new-lease-on-life/article_87440566-f2a0-11e8-9dce-73b7feb57639.html

Patton, A. (2022, June 24). UK PM champions social impact bonds – but are claims in Big Society Capital report justified? *Pioneers Post.* https://www.pioneerspost.com/news-views/20220624/bsc-outcomes-contracts

Patton, M. Q. (1990). *Qualitative evaluation and research methods.* Thousand Oaks, CA: Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods (3rd ed.).* Thousand Oaks, CA: Sage.

Paxon Group. (2020). *Silver Chain - Evaluation of the Silver Chain Community Palliative Care Service Western Sydney – Social Impact Investment.* https://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/files/Silver-Chain-Implementation-Evaluation-Report.pdf

Pearson A. (2004). Balancing the evidence: incorporating the synthesis of qualitative data into systematic reviews. *JBI Reports 2*(2), 45–64. http://dx.doi.org/10.1111/j.1479-6988.2004.00008.x

Picker, V., Carter, E., Airoldi, M., Ronicle, J., Wooldridge, R., Llewellyn, J., Monk, L., Stone, S., Gibson, M., Rosenbach, F., & Hameed, T. (2021). Social Outcomes Contracting (SOC) in Social Programmes and Public Services: A Mixed-Methods Systematic Review Protocol. *Social Science Protocols. 4*(2021). https://doi.org/10.7565/ssp.v4.5430

Plewis, I. (2000). Educational Inequalities and Education Action Zones. In C. Pantazis and D. Gordon (Eds.), *Tackling Inequalities: Where Are We Now and What Can Be Done* (pp. 86–100). Policy Press.

Pocock, J. (1971). *Politics, Language and Time: Essays on Political Thought and History.* University of Chicago Press.

Pollitt, C., & Bouckaert, G. (2011). *Continuity and change in public policy and management.* London, UK: Edward Elgar Publishing.

Pollitt, C., & Bouckaert, G. (2017). *Public management reform: A comparative analysis - into the age of austerity (4th ed.).* Oxford University Press.

Pollock, T., & Rindova, V. (2003). Media legitimation effects in the market for initial public offerings. *Academy of Management Journal, 46*(5), 631-642. http://dx.doi.org/10.2307/30040654

Ponzi, L., & Koenig, M. (2002). Knowledge management: another management fad? *Information Research, 8*(1).

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme.* https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf

Post Reporters (2022, July 04). Danny Kruger: 'The UK is still the best country for social investment'. *Pioneers Post.* https://www.pioneerspost.com/news-views/20220704/danny-kruger-the-uk-still-the-best-country-social-investment

Potts, J., & Kastelle, T. (2010). Public sector innovation research: What's next? *Innovation, 12*(2), 122-137. http://dx.doi.org/10.5172/impp.12.2.122

Priyanta, S., & Trisna, N. (2019). Social Network Analysis of Twitter to Identify Issuer of Topic using PageRank. *International Journal of Advanced Computer Science and Applications, 10*(1), 2019. http://dx.doi.org/10.14569/IJACSA.2019.0100113

Quirke, L. (2013). Rogue resistance: Sidestepping isomorphic pressures in a patchy institutional field. *Organization Studies, 234*(11), 1675-1699. http://dx.doi.org/10.1177/0170840613483815

Rao, H., Greve, H.R., and Davis, G.F. (2001). Fool's gold: Social proof in the initiation and abandonment of coverage by Wall Street analysts. *Administrative Science Quarterly, 46*(3). http://dx.doi.org/10.2307/3094873

Raven, M., Niedzwiecki, M., &Kushel, M. (2020). A randomized trial of permanent supportive housing for chronically homeless persons with high use of publicly funded

services. *Health Services Research, 55*(S2), 797-806. https://doi.org/10.1111/1475-6773.13553

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems, 89*(2015), 14–46. http://dx.doi.org/10.1016/j.knosys.2015.06.015

Reeder, N., Colantonio, A., Loder, J., Rocyn Jones, G. (2015). Measuring impact in impact investing: an analysis of the predominant strength that is also its greatest weakness. *Journal of Sustainable Finance & Investment, 5*(3), 136-154. http://dx.doi.org/10.1080/20430795.2015.1063977

Reeder, N., Rocyn Jones, G., Loder, J., Colantonio, A. (2014). *Measuring Impact: Preliminary insights from interviews with impact investors.* London School of Economics and Political Science. https://sosyaletkianalizi.com/wp-content/uploads/2017/11/measuring_impact_interviews_with_impact_investors.pdf

Rees, R., Oliver, K., Woodman J., & Thomas J. (2009). *Children's views about obesity, body size, shape and weight. A systematic review*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Resnik, D. B. (2000). Financial interests and research bias. *Perspectives on Science, 8*(3), 255-285. http://dx.doi.org/10.1162/106361400750340497

Rimer, B. K.; Glanz, K. (2005). *Theory at a glance: a guide for health promotion practice (2nd ed).* U.S. Department of Health and Human Services. https://www.sbccimplementationkits.org/demandrmnch/wp-content/uploads/2014/02/Theory-at-a-Glance-A-Guide-For-Health-Promotion-Practice.pdf

Ritchie, J., & Spencer, L. (1994) Qualitative Data Analysis for Applied Policy Research. In: A Bryman., & R Burgess (Eds.), *Analyzing Qualitative Data* (pp. 173-194). Routledge.

Roberts, L & Cameron, G. (2014). *Evaluation of the Essex Multi Systemic Therapy Social Impact Bond. Interim Evaluation Report, OPM, UK.* https://traverse.ltd/application/files/9515/2285/2105/Interim-report-Essex-MST-SIB-Evaluation.pdf

Roberts, L., & Cameron, G. (2015). *Evaluation of the Essex MultiSystemic Therapy Social Impact Bond - Interim evaluation report.* https://traverse.ltd/application/files/9515/2285/2105/Interim-report-Essex-MST-SIB-Evaluation.pdf

Rochat, Y. (2014). *Character networks and centrality* [Doctoral dissteration, University of Lausanne]. https://serval.unil.ch/resource/serval:BIB_663137B68131.P001/REF

Rogers, E. M. (1962). *Diffusion of innovation.* New York: The Free Press.

Rogers, E. M. (1983). *Diffusion of innovation (2nd ed.)*. New York: The Free Press.

Rogers, E. (1995). *Diffusion of Innovations: Fourth Edition.* New York, USA: The Free Press.

Rogers, E.M. (2003). *Diffusion of innovations (5th ed.).* New York: The Free Press.

Rogers, E., & Shoemaker, F. (1971). *Communication of Innovations: A Cross-Cultural Approach.* New York, THe Free Press.

Ronicle, J & Stanworth, N. (2015). *Ways to Wellness Social Impact Bond: The UK's First Health SIB: Deep Dive Report, Commissioning Better Outcomes Evaluation.* https://golab.bsg.ox.ac.uk/documents/Ways_to_Wellness_Deep_Dive_Report_1_Final_High_Res.pdf

Ronicle, J., Fox, T., & Stanworth, N. (2016). *Commissioning Better Outcomes Fund Evaluation - Update Report. Ecorys.*

https://www.tnlcommunityfund.org.uk/media/research-documents/social-investment/CBO-Update

Ronicle, J., Fraser, A., Tan, S., & Erskine, C. (2017). *The LOUD SIB Model: The four factors that determine whether a social impact bond is launched.* https://golab.bsg.ox.ac.uk/documents/loud_sib_model.pdf

Ronicle, J., Stanworth, N., Hickman, E., Fox, T. (2014). *Social Impact Bonds: The State of Play (Full Report), Commissioning Better Outcomes Evaluation.* https://www.biglotteryfund.org.uk/research/social-investment/publications

Rosenback, F., & Carter, E. (2020). *Kirklees Integrated Support Service and Better Outcomes Partnership: The first report from a longitudinal evaluation of a Life Chances Fund impact bond.* https://golab.bsg.ox.ac.uk/documents/Kirklees_Full_Report_.pdf

Rotheroe (2014). *Lessons and opportunities: perspectives from providers of social impact bonds, New Philanthropy Capital (NPC).* http://www.thinknpc.org/publications/lessons-and-opportunities/

Røvik, K, A. (1998). *Moderne Organisasjoner: Trender i Organisasjonstenkninger Ved Tusenårsskiftet.* Norge: Fagbokforlaget.

Roy, M.J., McHugh, N., & Sinclair, S. (2017) Social Impact Bonds – Evidence Based Policy or Ideology? In: Greve, B. (Ed.), *Handbook of Social Policy Evaluation*. Edward Elgar Publishing.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688–701. https://doi.org/10.1037/h0037350

Rudd, T., Nicoletti, E., Misner, K., Bonsu, J. (2013). *Financing Promising Evidence-Based Programs: Early Lessons from the New York City Social Impact Bond, MDRC.* http://files.eric.ed.gov/fulltext/ED545464.pdf

Ruef, M., & Scott, R. (1998). A multidimensional model of organizational legitimacy: Hospital survival in changing institutional environments. *Administrative Science Quarterly, 43*(4), 877-904. http://dx.doi.org/10.2307/2393619

Rutter, H., Savona, N., Glonti, K., Bibby, J., Cummins, S., Finegood, T., Greaves, F., Harper, L., Hawe, P., Moore, L., Petticrew, M., Rehfuess, E., Shiell, A., Thomas, J., White, M. (2017). The need for a complex systems model of evidence for public health. *The Lancet, 390(10112)*, 2602. https://doi.org/10.1016/S0140-6736(17)31267-9.

Ryan, B., & Gross, N. (1943). The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociology,* 13, 273-285.

Sabatier, P. A., & Weible, C. M. (2014). *Theories of the policy process (P. A. Sabatier & C. M. Weible Eds. 3rd ed.).* Boulder, CO: Westview Press.

Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal, 312(*7023), 71-72. http://dx.doi.org/10.1136/bmj.312.7023.71

Salge, TO., Kohli, R., & Barrett, M. (2015). Investing in information systems: On the behavioral and institutional search mechanisms underpinning hospitals' IS investment decisions. *MIS Quarterly 39*(1): 61–89. http://dx.doi.org/10.25300/MISQ/2015/39.1.04

Salis, S., Wishart, R., (2018). *Evaluation of the Innovation Fund pilot Quantitative assessment of impact and social return on investment.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/737021/evaluation-of-the-innovation-fund-pilot-quantitative-assessment-of-impact-and-social-return-on-investment.pdf

Saltman, K. (2017). The Promise and Realities of Pay for Success/Social Impact Bonds. *Education policy analysis archives, 25*(2017), 59. http://dx.doi.org/10.14507/epaa.25.2640

Sanderson, I. (2002). Making Sense of 'What Works': Evidence Based Policy Making as Instrumental Rationality? *Public Policy & Administration, 17*(3), 61-75. http://dx.doi.org/10.1177/095207670201700305

Sandstrom, H., & Greenberg, E. (2016). *PFS + ECE: Using Data to Inform Decision making Pay for Success Early Childhood Education Toolkit Report #2.* https://pfs.urban.org/ece-toolkit-usingdata

Sato, K., Mano, T., Iwata, A. et al. Need of care in interpreting Google Trends-based COVID-19 infodemiological study results: potential risk of false-positivity. *BMC Medical Research Methodology 147(*2021). https://doi.org/10.1186/s12874-021-01338-2

Schneiberg, M., & Soule, S. A. (2005). Institutionalization as a contested, multilevel process: The case of rate regulation in American fire insurance. In G. Davis, D. McAdam, W. R. Scott, & M. Zald (Eds.), *Social movements and organization theory* (pp. 122–160). Cambridge, U.K.: Cambridge University Press.

Schon, D. (1971). *Beyond the stable state.* New York: Norton.

Schuck, P. (2014). *Why Government fails so often.* Princeton University Press, Princeton, New Jersey.

Schuster, J., & Kolleck, N. (2020). The Global Diffusion of Social Innovations – An Analysis of Twitter Communication Networks Related to Inclusive Education. *Frontiers in Education, 5*(2020). http://dx.doi.org/10.3389/feduc.2020.492010

Schwartz, G., Christensen, T., & Zhu, X. (2022). Bounded Rationality, Satisficing, Artificial Intelligence, and Decision-Making in Public Organizations: The Contributions of Herbert Simon. *Public Administration Review, 82*(5), 902-904. https://doi.org/10.1111/puar.13540

Scott, R., & Davis, G. (2007). *Organizations and Organizing: Rational, Natural and Open System Perspectives.* Routledge.

Scott, W. (1998). *Organizations: Rational, natural and open systems.* Englewood Cliffs, NJ: Prentice-Hall.

Scott, W. (2013). *Institutions and organizations: Ideas, interests and identities (4th ed.).* Thousand Oaks, CA: Sage.

Scott, W. R. (1987). The adolescence of institutional theory. *Administrative Science Quarterly, 32*(4), 493-511. http://dx.doi.org/10.2307/2392880

Scott, W. R. (1995). *Institutions and organizations.* Thousand Oaks, CA: Sage.

Scott, W. R. (2001). *Institutions and Organizations.* Thousand Oaks: Sage Publications.

Scott, W. R. (2008). *Institutions and Organizations, 3rd Edition.* Thousand Oaks, CA: Sage.

Scott, W. R., Ruef, M., Mendel, P., & Caronna, C. (2000). *Institutional change and healthcare organizations: From professional dominance to managed care.* Chicago: University of Chicago Press.

Shiell, A., Hawe, P., & Gold, L. (2008). Complex interventions or complex systems? Implications for health economic evaluation. *BMJ : British Medical Journal, 336*(7656), 1281–1283. http://doi.org/10.1136/bmj.39569.510521.AD.

Shipilov, A., Greve, H., & Rowley, T. (2010). When do interlocks matter? Institutional logics and the diffusion of multiple corporate governance practices. *Academy of Management Journal, 53*(4), 846-864. http://dx.doi.org/10.5465/amj.2010.52814614

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129-138. http://dx.doi.org/10.1037/h0042769

Sinclair, S., McHugh, N., & Roy, M. (2019). Social innovation, financialisation and commodification: a critique of social impact bonds, *Journal of Economic Policy Reform, 2*4(1), 11-27. http://dx.doi.org/10.1080/17487870.2019.1571415

Sinclair, S., McHugh, N., Huckfield, L., Roy, M. J., & Donaldson, C. (2014). Social Impact Bonds: Shifting the Boundaries of Citizenship. *Social Policy Review 26*, 2014, 119–136. http://dx.doi.org/10.1332/policypress/9781447315568.003.0007

Singh, J., Tucker, D., & House, R. (1986). Organizational legitimacy and the libaility of newness. A*dministrative Science Quarterly, 31*(2), 171-193.

Sinnenberg, L., Buttenheim, A., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. (2017). Twitter as a Tool for Health Research: A Systematic Review. *American Journal of Public Health, 107*(1), e1–e8. http://dx.doi.org/10.2105/AJPH.2016.303512

Slavin R. (1995). Best evidence synthesis: an intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology, 48*(1), 9–18. http://dx.doi.org/10.1016/0895-4356(94)00097-A

Small, H. G. (1978). Cited documents as concept symbols. *Social Studies of Science, 8*(3), 327-340. http://dx.doi.org/10.1177/030631277800800305

Smeets, D (2017). Collaborative Learning Processes in Social Impact Bonds: A Case Study from the Netherlands. *Journal of Social Entrepreneurship, 8*(1), 67-87. http://dx.doi.org/10.1080/19420676.2017.1299034.

Smith, K. (2013a). *Beyond evidence based policy in public health: The interplay of ideas.* London: Springer.

Smith, K. E. (2013b). Institutional filters: the translation and re-circulation of ideas about health inequalities within policy. *Policy and Politics, 41*(1), 81-100. http://dx.doi.org/10.1332/030557312X655413

Smullen, A. (2010). *Translating Agency Reform*. [Doctoral dissertation, Erasmus University Rotterdam.https://repub.eur.nl/pub/10428/THESIS%20AMANDA%20SMULLEN.pdf

Sniazhko, Sniazhana. (2019). Uncertainty in decision-making: A review of the international business literature. *Cogent Business & Management, 6*(1), 1650692. http://dx.doi.org/10.1080/23311975.2019.1650692

Social Finance. (2011). *Social Impact Bonds The One* Service. One year on.* https://www.socialfinance.org.uk/sites/default/files/publications/sf_peterborough_one-_year_on.pdf

Sowles, S., McLeary, M., Opticana, A., Cahna, E., Krauss, M., Fitzsimmons-Craft, E., Wilfleya, D., Cavazos-Rehga. (2018). A content analysis of an online pro-eating disorder community on Reddit. *Body Image, 24*(2018), 137-144. https://doi.org/10.1016/j.bodyim.2018.01.001

Spann, R. (1981). Fashions and Fantasies in Public Administration. *Australian Journal of Public Administration, 40*(1), 12-25. http://dx.doi.org/10.1111/j.1467-8500.1981.tb00481.x

Stanley JC. Controlled experimentation in the classroom. *The Journal of Experimental Education, 25*(3), 195–201. http://dx.doi.org/10.1080/00220973.1957.11010571

Stanworth, N. (2020). *Zero HIV Social Impact Bond - Commissioning Better Outcomes Fund Evaluation: In depth review.* https://golab.bsg.ox.ac.uk/documents/EJAF-Zero-HIV-in-depth-review.pdf

Stanworth, N., & Hickman, E. (2022). *The value created by social outcomes contracts in the UK. ATQ Consultants.* https://bigsocietycapital.fra1.cdn.digitaloceanspaces.com/media/documents/ATQ_SOC_Social_value_report.pdf

Statista (2021a, January 27). *Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019.* Retrieved January 12, 2023, from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Statista. (2021b, November). *Number of internet users worldwide from 2005 to 2021(in millions).* Retrieved January 12, 2023, from https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/

Staw, B. M., Sandelands, L. E., & Dutton, J. (1981). Threat-rigidity effects in organizational be multi-level analysis. *Administrative Science Quarterly, 26*(4), 501-524.

Stone, D. (2001) *Bridging research and policy.* Warwick: Warwick University.

Strang, D., & Soule, S. (1998). Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual Review of Sociology, 24*(1), 265-290. http://dx.doi.org/10.1146/annurev.soc.24.1.265

Straus, S., Glasziou, P., Richardson, W. S., & Haynes, R. B. (2018). *Evidence-Based Medicine: How to Practice and Teach EBM.* Elsevier Health Sciences.

Strickland, P. (2010). *Social Impact Bonds – The Pilot at Peterborough Prison.* House of Commons Library Briefing. London: HMSO. http://www.parliament.uk/briefingpapers/commons/lib/research/briefings/SNHA-05758.pdf

Struewer, B., & Moehrle, C. (2015, December 14). NexThought Monday: Pay for Success … With an Important Twist. *Next Billion.* https://nextbillion.net/nexthought-monday-pay-for-success-with-an-important-twist/

Suchman, M. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Management Review, 20*(3), 571-610. http://dx.doi.org/10.2307/258788

Suri H. (2000). A critique of contemporary methods of research synthesis. *Post-Script 1*(2000), 49-55.

Suri, H. (2011). Purposeful sampling in qualitative research synthesis. *Qualitative Research Journal, 11*(2), 63-75. http://dx.doi.org/10.3316/QRJ1102063

Talbot, C., O'Dwyer, S., Clare, L., Heaton, J., Anderson, J. (2020). How people with dementia use twitter: A qualitative analysis. *Computers in Human Behavior, 102*(2020), 112-119. http://dx.doi.org/10.1016/j.chb.2019.08.005

Taleb, Nassim N. (2010). *The Black Swan: The Impact of the Highly Improbable, Second ed.* London: Penguin Books.

Tan, S., Fraser, A., Giacomantonio, C., Kruithof, K., Sim, M., Lagarde, M., Disley, E., Rubin, J., Mays, N. (2015). *An evaluation of Social Impact Bonds in health and social care: interim report, Policy Innovation Research Unit (PIRU) (Publication 2015 – 12).* Available at: http://researchonline.lshtm.ac.uk/2391521/

Tanninen, K., Jantunen, A., & Saksa, J. (2008). Adoption of administrative innovation within organization - An empirical study of TQM metamorphosis. *International Journal of Innovation and Technology Management 05*(03), 321-340. http://dx.doi.org/10.1142/S0219877008001412

Taylor, D., & and Balloch, S (eds) (2005). *The Politics of Evaluation: Participation and Policy Implementation.* Bristol: Policy Press.

Teece, D. (1977). Technology transfer by multinational firms: The resource costs of transferring technological know-how. *The Economic Journal, 87*(346), 242-261. http://dx.doi.org/10.2307/2232084

Teece, D. 1980. The diffusion of an administrative innovation. *Management Science, 26*(5), 464–470. http://dx.doi.org/10.1287/mnsc.26.5.464

Thelwall, M &., Thelwall, S. (2020). A thematic analysis of highly retweeted early COVID - 19 tweets: Consensus, information, dissent, and lockdown life. *ASLIB Journal of*

Information Management, 72(6), 945-962. http://dx.doi.org/10.1108/AJIM-05-2020-0134

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology. 8*(45). https://doi.org/10.1186/1471-2288-8-45

Thomas. (1999). *Fashions in Management Research. An Empirical Analysis (First Edition).* Routledge.

Thompson, J. (1967). *Organization in Action: Social Science Basis of Ad ministrative Theory.* New York: Mc Graw-Hill Book Co.

Thomson, H., Thomas, S., Sellstrom, E., & Petticrew, M. (2013). Housing improvements for health and associated socio-economic outcomes. *The Cochrane database of systematic reviews, (2)*, CD008657. https://doi.org/10.1002/14651858.CD008657.pub2

Thorne, S., Jensen, L., Kearney, MH., Noblit, G., Sandelowski, M. (2004). Qualitative meta-synthesis: reflections on methodological orientation and ideological agenda. *Qualitative Health Research, 14*(10), 1342-1365. https://doi.org/10.1177/1049732304269888

Thornton, P. H. (2004). *Markets from culture: Institutional logics and organizational decisions in higher education publishing.* Stanford, CA: Stanford University Press.

Thornton, P., Ocasio, W., & Lounsbury, M. (2012). *The Institutional Logics Perspective: A New Approach to Culture, Structure and Process.* USA: Oxford University Press.

Tinati R., Halford, S., Carr L., & Pope, C/ (2014). Big data: Methodological challenges and approaches for sociological analysis. *Sociology 48*(4): 663–681. http://dx.doi.org/10.1177/0038038513511561

Tolbert, P., & Zucker, L. (1983). Institutional sources of change in the formal structure of organizations: The diffusion of civil service reform, 1880-1935. *Administrative Science Quarterly, 28*(1), 22-39.

Tomkinson, E. (2012). *Perspectives from the Social Finance Forum 2012: An Australian Snapshot - Social Impact Bonds, the Centre for Social Impact (CSI).* http://www.csi.edu.au/research/project/social-impact-bonds-australian-snapshot/

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care, 19*(6), 349–357. http://doi.org/10.1093/intqhc/mzm042

Torres Hernandez, V. (2008). Making good on the promise of international law: the convention on the rights of persons with disabilities and inclusive education in China and India. *Washington International Law Journal, 17*(2), 497–527.

Tse, A., & Warner, M. (2018). The razor's edge: Social impact bonds and the financialization of early childhood services. *Journal of Urban Affairs, 42*(6), 816-832, http://dx.doi.org/10.1080/07352166.2018.1465347

Tsoukas, H., & Knudsen, C. (2005). *The Oxford Handbook of Organization Theory: Meta-theoretical perspectives.* Oxford University Press.

Tushman, M. L. and Anderson, P. D. (1986). Technology discontinuities and organizational environments. *Administrative Science Quarterly, 31*(3), 439–464

UNDP (2016). *Social and development impact bonds (results-based financing).* www.undp.org/content/sdfinance/en/home/solutions/social-development-impactbonds.html.

van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., Kramer, B., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Yongchao, M., Fang, Q., Hindriks, S., Tummers, L., & Oberski, D. (2021). An open source machine

learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence 3*(2), 125–133. http://dx.doi.org/10.1038/s42256-020-00287-7

Van de Ven, A. H. (1986). Central problems in the management of innovation. *Management Science, 32*(5), 590-607. http://dx.doi.org/10.1287/mnsc.32.5.590

Van de Ven, A., Ganco, M & Hinings, B. (2013). Returning to the Frontier of Contingency Theory of Organizational and Institutional Designs. *The Academy of Management Annals, 7*(1), 391-438. http://dx.doi.org/10.5465/19416520.2013.774981

Van der Nagel, E., & Frith, J. (2015). Anonymity, pseudonymity, and the agency of online identity: Examining the social practices of r/Gonewild. *First Monday, 20*(3). http://dx.doi.org/10.5210/fm.v20i3.5615

Van Es, B., Houben, O., & Nijeholt, M. (2016). Social Impact Bonds: An Innovative Way to Combat Youth Unemployment. *Journal of Positive Management, 7*(1), 55-56. https://doi.org/10.12775/JPM.2016.003

Vandenbroucke, J., Broadbent, A., Pearce, N. (2016). Causality and causal inference in epidemiology: the need for a pluralistic approach. *International Journal of Epidemiology, 45*(6), 1776–1786. https://doi.org/10.1093/ije/dyv341.

Vardi, E., Muchnik, L., Conway, A., & Breakstone, M. (2021, April). WikiShark: An Online Tool for Analyzing Wikipedia Traffic and Trends. Companion Proceedings of the Web Conference, 2021, 558-571. http://dx.doi.org/10.1145/3442442.3452341

Vecchi, V., & Casalini, F. (2019). Is a Social Empowerment of PPP for Instrastructure Delivery Possible? Lessons from Social Impact Bonds. *Annals of Public and Cooperative Economics, 90*(2), 353-369. https://doi.org/10.1111/apce.12230

Verboom, B., Montgomery, P. & Bennett, S. What factors affect evidence-informed policymaking in public health? Protocol for a systematic review of qualitative evidence using thematic synthesis. *Systematic Reviews 5*(61), 2016. https://doi.org/10.1186/s13643-016-0240-6

Vivona, R., Demircioglu, M., & Raghavan, A. (2020). Innovation and Innovativeness for the Public Servant of the Future: What, Why, How, Where, and When. In: *The Palgrave Handbook of the Public Servant* (pp. 1-22), Springer Nature.

Wainer, A. (2018). *Investing in Maternal and Child Health: Development Impact Bonds Potential and Early Learning.* https://golab.bsg.ox.ac.uk/documents/Save_the_Children_2018_Investing_in_Maternal_and_Child_Health_-_DIBs.pdf

Walitzer KS, Dermen KH, Barrick C, Shyhalla K. Modeling the Innovation-Decision Process: Dissemination and Adoption of a Motivational Interviewing Preparatory Procedure In Addiction Outpatient Clinics. *Journal of Substance Abuse Treatment, 57*(2015), 18-29. http://dx.doi.org/10.1016/j.jsat.2015.04.003

Walker, R. (2003). Evidence on the management of public services innovation, *Public Money and Management 23*(2), 93-102. http://dx.doi.org/10.1080/09540962.2003.10874830

Walker, K., F. Schlosser, and D. L. Deephouse. 2014. Organizational Ingenuity and the Paradox of Embedded Agency: The Case of the Embryonic Ontario Solar Energy Industry. *Organization Studies, 35*(4), 613–634. http://dx.doi.org/10.1177/0170840613517599

Wallace, A., & Roxas, B. (2014). *Evaluation of the Newpin SBB Program. Implementation Report.* https://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/files/Newpin-Evaluation-Implementation-Report.pdf

Wallace, A., Hutley, N., Tomiczek, C., & Batchen, A. (2016). *Evaluation of the Newpin SBB Program – Interim Evaluation Report.* https://www.osii.nsw.gov.au/assets/Uploads/Interim-Report-Evaluation-of-the-Newpin-SBB-Program-October-2016.pdf

Wallace, A., Wise, C., & Griffiths, C. (2018). *Newpin – Second Interim Evaluation Report.* https://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/Newpin-Second-Interim-Evaluation-Report-2018-Final.pdf

Walsh, K., Bieretz, B., Walker, K., & Mitra-Majumdar, M. (2017). *Why Governments and Investors Choose Pay for Success: Perspectives on Motivations and Risks.* https://policycommons.net/artifacts/631332/why-governments-and-investors-choose-pay-for-success/1612656/

Wang, B. and Zhuang, J. (2017). Crisis information distribution on Twitter: A content analysis of tweets during Hurricane Sandy. *Natural Hazards, 89*(1), 161-181. http://dx.doi.org/10.1007/s11069-017-2960-x

Warner, M. (2013). Private finance for public goods: social impact bonds. *Journal of Economic Policy Reform, 16*(4), 303-319. Available at: http://dx.doi.org/10.1080/17487870.2013.835727

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge University Press.

Webb, H., Jirotka, M., Stahl, BC., Housley, W., Edwards, A., Williams, M., Procter, R., Rana, O., & Burnap, P. (2017). The ethical challenges of publishing Twitter data for research dissemination. In *Proceedings of the 2017 ACM on Web Science Conference*. http://dx.doi.org/10.1145/3091478.3091489

Webb, S. A. (2001). Some considerations on the validity of evidence-based practice in social work. *British Journal of Social Work, 31*(1), 57-79. http://dx.doi.org/10.1093/bjsw/31.1.57

Weber, K., Davis, G., & Lounsbury, M. (2009). Policy as myth and ceremony? The global spread of stock exchanges, 1980-2005. *Academy of Management Journal, 52*(6), 1319-1347.

Weber, M. (1947). *The Theory of Social and Economic Organization (trans. T. Parsons).* New York: Free Press.

Wejnert B. (2002). Integrating Models of Diffusion of Innovations: A Conceptual Framework. *Annual Review of Sociology, 28*(1), 297–326. http://dx.doi.org/10.1146/annurev.soc.28.110601.141051

Welch, C., & Mihalcea, R. (2017). Targeted sentiment to understand student comments. In *Proceedings of the International Conference on Computational Linguistics (COLING 2016)* (pp. 2471–2481). https://aclanthology.org/C16-1233.pdf

Westphal, J., & Zajac, E. (2001). Decoupling policy from practice: The case of stock repurchase programs. *Administrative Science Quarterly 46*(2), 202-228. http://dx.doi.org/10.2307/2667086

Westphal, James D., Ranjay Gulati, and Steve Shortell. (1997). Customization or conformity? An institutional and network perspective on the content and consequences of TQM adoption. *Administrative Science Quarterly, 42*(2), 366–94. http://dx.doi.org/10.2307/2393924

Whewell, W. (1840, 1847). *The philosophy of the inductive sciences: Founded upon their history.* London: J.W. Parker.

Williams (2019). From Visions of Promise to Signs of Struggling: Exploring Social Impact Bonds and the Funding of Social Services in Canada, the US and the UK. https://fss.info.yorku.ca/files/2019/05/Williams-2019-Final-Report-May-1.pdf?x25736

Williams, M., Burnap, P., & Sloan, L. (2017). Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation. *Sociology, 51*(6), 1149–1168. http://dx.doi.org/10.1177/0038038517708140

Wilson, K. E. (2014). *Social Investment: New Investment Approaches for Addressing Social and Economic Challenges.* Paris: OECD Library.

Wiltz, T. (2016). *States, Localities Try Creative Financing to Help the Homeless.* https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2016/02/22/states-localities-try-creative-financing-to-help-the-homeless

Windrum, P., Koch, P. (Eds.). (2008). *Public Sector Services – Entrepreneurship, Creativity and Management*. Cheltenham, UK/Northampton: Edward Elgar.

Wise, P., Formosa, J., Wiseman, A., Grisdale, J., Farmer, J., & Fung, C. (2022). *Evaluation of the Resolve Program – Final Interim Report 2022.* https://www.osii.nsw.gov.au/assets/office-of-social-impact-investment/files/ESA31417-Resolve-Social-Benefit-Bond-Interim-Report-13-July-2022.pdf

Wooldridge, Stanworth & Ronicle (2019). *A study into the challenges and benefits of commissioning Social Impact Bonds in the UK, and the potential for replication and scaling: final report.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957374/A_study_into_the_challenges_and_benefits_of_the_SIB_commissioning_process._Final_Report_V2.pdf

Wynn, R., Oyeyemi, O., Johnson, J., & Gabarron, E. (2017). Tweets are not always supportive of patients with mental disorders. *International Journal of Integrated Care, 17*(3), A149. http://dx.doi.org/10.5334/ijic.3261

Yaffe, J., Montgomery, P., Hopewell, S., Shepard, L. (2012). Empty Reviews: A Description and Consideration of Cochrane Systematic Reviews with No Included Studies. *PLoS ONE 7*(5), e36626. https://doi.org/10.1371/journal.pone.0036626

Yang, Q., Tufts, C., Ungar, L., Guntuku, S., & Merchant, R. (2018). To Retweet or Not to Retweet: Understanding What Features of Cardiovascular Tweets Influence their Retransmission. *Journal of Health Communication, 23*(12): 1026–1035. http://dx.doi.org/10.1080/10810730.2018.1540671

Young, P. (2015). The Daily SIB Newsletter - October 8 2015. *Social Impact Bond: The Daily SIB Newsletter.* http://socialimpactbondnews.com/october-8-2015/

Young, P. (2016). The Daily SIB Newsletter - July 07 2016. *Social Impact Bond: The Daily SIB Newsletter.* http://socialimpactbondnews.com/july-07-2016/

Yousafzai, M. (2020). Malala: We need a reset, not a recovery. *Evoke.* https://www.evoke.org/articles/october-2020/how-i-hope-we-rebuild-the-world-after-covid-19-malala

Zaltman, G., Duncan, R., & Holbeck, J. (1973). Innovation and organizations. New York: Wiley.

Zamani, M, Rabbani, F, Horicsányi, A, et al. (2019) Differences in structure and dynamics of networks retrieved from dark and public web forums. *Physica A: Statistical Mechanics and Its Applications*, *525(*2019), 326–336. http://dx.doi.org/10.1016/j.physa.2019.03.048

Zhai, Y., Ding, Y., & Wang, F. (2017). Measuring the diffusion of an innovation: A citation analysis. *Journal of the Association for Information Science and Technology, 69*(3):368–379, 2018. http://dx.doi.org/10.1002/asi.23898

Zhang, G., Ding, Y., & Milojevic´, S. (2013). Citation content analysis (cca): A framework for syntactic and semantic analysis of citation content. *Journal of the American Society for Information Science and Technology, 64*(7), 1490–1503. http://dx.doi.org/10.1002/asi.22850

Zhang, J., & Luo, Y. (2017). Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network. *Advances in Intelligent Systems Research, 132*(2017). http://dx.doi.org/10.2991/msam-17.2017.68

Zimmerman, M., & Zeitz, G. (2002). Beyond survival: Achieving new venture growth by building legitimacy. Academy of Management Review, 27(3), 414-431. http://dx.doi.org/10.2307/4134387

Zucker, L. (1983). Organizations as Institutions. In Bacharach, S. (Ed.), *Perspectives in Organizational Sociology: Theory and Research*. ASA Series.

Zucker, L. (1991). Postscript: Microfoundations of institutional thought. In W.W. Powell., & P.J. DiMaggio (Eds.), *The New Institutionalism in Organizational Analysis* (pp. 103-106). Chicago: University of Chicago Press.

Zwahlen, M., & Salanti, G. (2018). Causal inference from experiment and observation. *Evidence-Based Mental Health 21*(1). http://dx.doi.org/10.1136/eb-2017-102859

Appendices

Appendix A: Example Search Terms

SCOPUS 1990–present

1.TITLE-ABS-KEY ("outcome*-base* contract*")
2.TITLE-ABS-KEY ("outcome*-contract*")
3.TITLE-ABS-KEY ("Result*-base* contract*")
4.TITLE-ABS-KEY ("outcome*-base* commissioning")
5.TITLE-ABS-KEY ("Outcome*-base* procurement")
6.TITLE-ABS-KEY ("social outcome*" W/5 "contract*")
7.TITLE-ABS-KEY ("social outcome*" AND "contract*")
8.TITLE-ABS-KEY ("performance-base* incentive*")
9.TITLE-ABS-KEY ("Result*-base* financ*")
10.TITLE-ABS-KEY ("Performance-base* financ*")
11.TITLE-ABS-KEY ("Sustainability-linked loan*")
12.TITLE-ABS-KEY ("Performance-base* contract*")
13.TITLE-ABS-KEY ("Performance-base* aid")
14.TITLE-ABS-KEY ("output-base* aid")
15.TITLE-ABS-KEY ("result*-base* aid")
16.TITLE-ABS-KEY ("Outcome*-base* financ*")
17.TITLE-ABS-KEY ("Performance base* transfer")
18.TITLE-ABS-KEY ("social" W/3 "public-private partnership*")
19.TITLE-ABS-KEY ("social" AND "outcome*" W/3 "public-private partnership*")
20.TITLE-ABS-KEY ("public-private mix")
21.TITLE-ABS-KEY ("social" W/3 "PPP")
22.TITLE-ABS-KEY ("social" AND "outcome*" W/3 "PPP")
23.TITLE-ABS-KEY ("contract* for health*")
24.TITLE-ABS-KEY ("payment-by-result*")
25.TITLE-ABS-KEY ("pay-by-result*")
26.TITLE-ABS-KEY ("Payment-by-results contract*")
27.TITLE-ABS-KEY ("pay*-by-outcome*")
28.TITLE-ABS-KEY ("Program-for-result*")
29.TITLE-ABS-KEY ("pay for success")
30.TITLE-ABS-KEY ("Payment-by-results contract*")
31.TITLE-ABS-KEY ("Pay*-for-outcome*")
32.TITLE-ABS-KEY ("Outcome-based payment*")
33.TITLE-ABS-KEY ("Outcome-base* financ*")
34.TITLE-ABS-KEY ("Social impact invest*")
35.TITLE-ABS-KEY ("Impact auction*")
36.TITLE-ABS-KEY ("impact bond" OR "impact bonds") OR TITLE-ABS-KEY ("social" W/4 ("impact bond" OR "impact bonds"))
37.TITLE-ABS-KEY ("social impact bond*")
38.TITLE-ABS-KEY ("Bonos de Impacto Social")
39.TITLE-ABS-KEY ("Development Impact Bond*")
40.TITLE-ABS-KEY ("Health Impact Bond*")
41.TITLE-ABS-KEY ("Social Impact Project*")
42.TITLE-ABS-KEY ("social impact partnership")

43.TITLE-ABS-KEY ("social impact" W/4 "partnership")
44.TITLE-ABS-KEY ("social success note*")
45.TITLE-ABS-KEY ("Contract" W/4 "service provider")
46.TITLE-ABS-KEY ("Alliance contract*")
47.TITLE-ABS-KEY ("performance-base*-grant*")
48.TITLE-ABS-KEY ("Outcome*-base*-grant*")
49.TITLE-ABS-KEY (“Targeted grant*”)
50.TITLE-ABS-KEY ("Disbursement-linked-indicator*")
51.TITLE-ABS-KEY ("Performance-base*-condition*")
52.Combined search 1 –51 (OR)
53.TITLE-ABS-KEY ("randomized controlled trial*")
54.TITLE-ABS-KEY ("randomised controlled trial*")
55.TITLE-ABS-KEY ("RCT*")
56.TITLE-ABS-KEY ("controlled clinical trial*")
57.TITLE-ABS-KEY ("interrupted time series analysis")
58.TITLE-ABS-KEY ("interrupted time series analyses")
59.TITLE-ABS-KEY ("randomis*" OR "randomiz*" OR "randomly")
60.TITLE-ABS-KEY ("pretest-posttest study" OR "pretesting" OR "pre-post tests" OR "quasi-experimental design" OR "quasi-experimental study" OR "quasi-experimental study design" OR "repeated measurement" OR "repeated measurements" OR "repeated measures" OR "time series")
61.TITLE-ABS-KEY ("meta analysis")
62.TITLE-ABS-KEY ("meta analyses")
63.TITLE-ABS-KEY (("systematic reviews"))
64.TITLE-ABS-KEY ("trial" OR "multicenter" OR "multi center" OR "multicentre" OR "multi centre")
65.TITLE-ABS-KEY ("evaluat*" OR "assess*" OR "impact" OR "measur*" OR "experience*" OR "perception*" OR "learning" OR "performance" OR "program*" OR "random*" OR "experiment*" OR "control*" OR "ebp" OR "evidence based practice" OR "outcome*")
66.TITLE-ABS-KEY ("intervention" OR "effect" OR "impact" OR "controlled" OR "control group" OR ("before" W/5 "after") OR ("pre" W/5 "post") OR ("pretest" OR "pre test")) AND TITLE-ABS-KEY ("posttest" OR "post test" OR "quasiexperiment*" OR "quasi experiment" OR "evaluat*" OR "time series" OR "time point" OR ("repeated measure" OR "repeated measurement" OR "repeated measurements" OR "repeated measures") OR "generalized estimating equation" OR "generalised estimating equation")
67.TITLE-ABS-KEY ("pragmatic clinical trial*")
68.TITLE-ABS-KEY ("non-randomized control* trial*")
69.TITLE-ABS-KEY ("non-randomised control* trial*")
70.TITLE-ABS-KEY ("controlled before-after studies")
71.TITLE-ABS-KEY ("cost*" OR "value for money" OR "VFM" OR "economic analysis" OR "economic analyses" OR "cost benefit analysis" OR "CBA" OR "transaction cost*")
72.TITLE-ABS-KEY ("interview*" OR "focus group*" OR "observation*" OR "case stud*" OR "reflective diar*")
73.Combined search 53 –72 (OR)
74.Combined search (52 AND 73)

Appendix B: Grey literature sources

The table below shows the number of results that were retrieved by manually searching the grey literature sources listed below. As with Appendix D, I present the results for a prior (2016) search versus an updated (2022) search, in order to show how the evidence base has developed, over time. In some instances, it was not possible to retrace/ locate all of the articles I located in 2016, thus resulting in a lower number of results in 2022. Overall, however, the number of grey literature source, identified from searching the websites listed below, grew drastically between 2016 and 2022 (from 513 sources to 1,835 sources).

#	Source	# of results (2016)	# of results (2022)
1	Google Scholar	207	350
2	Urban Institute	33	96
3	Social Finance	26	52
4	Government Performance Lab (Harvard)	25	19
5	Saïd Business School	19	10
6	Office of Social Impact Investment - NSW	18	64
7	Gov.uk	16	169
8	Global Impact Investing Network	10	63
9	Big Lottery Fund	9	122
10	Center for American Progress	8	59
11	Instiglio	7	11
12	Impact Investing Australia	6	11
13	National Council for Voluntary Organisations (NCVO)	6	2
14	RAND Europe	6	19
15	Stanford Social Innovation Review	5	9
16	Boston Consulting Group	4	9
17	Big Society Capital	4	20
18	Bridges Ventures	4	65
19	Carleton University	4	13
20	Deloitte	4	19
21	Rockefeller Foundation	4	10
22	Harvard University – Business School	3	20
23	New Philanthropy Capital	3	42
24	Canadian Centre for Policy Alternatives	2	19
25	Centre for Social Impact (CSI), NSW	2	21
26	Duke University	2	27
27	Global Social Impact Investment Steering Group	2	23

28	Government Outcomes Lab (GO Lab)	2	37
29	Government of the Netherlands	2	1
30	Mission Investors	2	15
31	OPM Group / Traverse	2	3
32	PwC	2	4
33	The Young Foundation	2	2
34	Third Sector Capital Partners	2	36
35	Venture Philanthropy Partners	2	3
36	World Economic Forum	2	4
37	2020 Public Services Trust	1	0
38	Acumen	1	1
39	Barclay's Wealth	1	1
40	Bill & Melinda Gates Foundation	1	0
41	Brookings Institute	1	63
42	Center for Global Development	1	28
43	Center for Law and Social Policy (CLASP)	1	7
44	Central European University	1	4
45	Cicero Consultancy	1	1
46	City of London	1	2
47	Corporation for National & Community Service	1	8
48	Criminal Justice Inspectorates	1	0
49	Early Intervention Foundation	1	5
50	European Services Strategy Unit	1	7
51	Evaluating Impact Investing (Jackson & Associates)	1	0
52	Federal Reserve Bank of Atlanta	1	3
53	General Assembly of Maryland	1	2
54	Hewlett Foundation	1	3
55	Impact Assets	1	0
56	Investing for Good	1	3
57	Istanbul Policy Center	1	0
58	Kauffman Foundation	1	1
59	KPMG	1	6
60	London.gov.uk	1	2
61	Manpower Demonstration Research Corporation (MDRC)	1	13
62	MaRS	1	9
63	McKinsey & Company	1	5
64	Milken Institute	1	14
65	Monitor Institute	1	0
66	National Australia Bank (NAB)	1	20
67	National Audit Office (NAO)	1	8
68	NHS Networks	1	1
69	Nonprofit Finance Fund	1	7
70	OECD	1	35
71	Omidyar Network	1	5

72	Parliament of NSW	1	31
73	Philanthropy Impact	1	8
74	PIRU	1	24
75	Policy and Politics (Journal Blog)	1	10
76	Princeton University	1	3
77	REES Centre, University of Oxford	1	2
78	Shelter SA	1	1
79	Social Impact Investment Taskforce	1	1
80	Social Market Foundation	1	10
81	Social Venture Technology Group	1	0
82	The Australian Sociological Association	1	0
83	The Centre for Social Justice	1	0
84	The Institute for Government	1	4
85	Tides Foundation	1	1
86	University of Leeds	1	4
87	Vera Institute of Justice	1	16
88	World Bank	1	3
89	Parliament.uk	1	4
	Total # of results:	513	1,835

Appendix C: Hand searching of journals

The table below shows the number of journal articles that were identified by hand searching journals (to account for papers that have not been properly indexed in databases). To show how the evidence base has developed, I present the results for a prior (2016) search versus an updated (2022) search. Note: in some instances, it was not possible to retrace/ locate the articles I located in 2016, thus resulting in a lower number of results in 2022.

Overall, however, the number of articles located by hand searching, grew significantly, from 49 to 121. Of the 121 articles, 68 were new articles which had not been picked up by the database searches or grey literature search. The high number of new articles demonstrates the importance of including hand searching as a component of the search strategy.

	Source	# of results (2016)	# of results (2022)
1	Journal of Sustainable Finance and Investment	8	3
2	Community Development	5	4
3	Journal of Social Entrepreneurship	3	9
4	Research in International Business and Finance	2	3
5	Journal of Positive Management	2	0
6	Entrepreneurship Research Journal	2	4
7	Accounting, Organizations and Society	1	2
8	Administrative Sciences	1	1
9	Business & Society	1	1
10	Canadian Journal of Nonprofit & Social Economy Research	1	1
11	Company and Securities Law Journal	1	0
12	Corporate Ownership and Control	1	2
13	Cultural Studies	1	2
14	European Scientific Journal	1	1
15	Financial Analysts Journal	1	2
16	Florida Tax Review	1	1
17	Fordham Law Review	1	1
18	Global Economics and Management Review	1	1
19	International Journal of Public Administration	1	4

20	International Journal of Public Policy	1	0
21	International Society for Third Sector Research	1	0
22	International Studies Review	1	0
23	Journal of Business Ethics	1	9
24	Journal of Comprehensive Health	1	0
25	Journal of Contemporary European Research	1	0
26	Journal of Law & Business*	1	0
27	Probation Journal	1	6
28	Procedia - Social and Behavioral Sciences	1	1
29	Research on Social Work Practice	1	1
30	Universal Journal of Accounting and Finance	1	1
31	University of Pennsylvania Journal of Business Law	1	0
32	University of Pittsburgh Law Review	1	1
33	VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations	1	14
34	Public Money & Management	0	1
35	Sustainability	0	1
36	Journal of Urban Affairs	0	
37	Journal of Economic Policy Reform	0	4
38	Journal of Comparative Policy Analysis	0	5
39	Research in International Business and Finance	0	5
40	Environment and Planning A-Economy and Space	0	4
41	Historical Social Research-Historische Sozialforschung	0	2
42	Public Administration Review	0	3
43	Brazilian Journal of Public Administration	0	1
44	International Journal of Public Sector Management	0	1
45	Nonprofit and Voluntary Sector Quarterly	0	2
46	Corporate Social Responsibility and Environmental Management	0	1
47	European Company and Financial Law Review	0	1
48	Nonprofit Policy Forum	0	15
	Total results	49	121

Appendix D: Experts contacted and sample email

Various experts (n = 54) were contacted to identify additional (published or unpublished) articles of relevance. The table below includes the details of the experts who had responded to the request at the time of submission. At the time of submission, the experts had suggested 53 articles, 14 of which had not been previously identified.

#	Name	Organisation
1	Professor Chris Fox	Manchester Metropolitan University
2	Dr Alec Fraser**	King's College London
3	Dr Kevin Albertson **	Manchester Metropolitan University
4	Dr Jonathan Kimmitt**	Newcastle University Business School
5	Dr Toby Lowe	Northumbria University
6	Annarita Trotta	Universita di Catanzaro
7	Manuel Wirth	University of Zurich
8	Jacob Broom	University of Western Australia
9	Helen Chiappini	G. d'Annunzio University of Chieti and Pescara
10	Samantha Iovan MPH	University of Michigan
11	Lisa Knoll	University of Hamburg
12	Paula Lantz**	University of Michigan
13	Mildred Warner	Cornell University
14	James Williams	York University
15	Juan David Rivera Acevedo	University of Erfurt, Germany
16	Irene Bengo	Politecnico di Milano
17	Christian Berndt	University of Zurich
18	Annette Boaz	University of London
19	Mario Calderini	Politecnico di Milano
20	Veronica Chiodo	Politecnico di Milano
21	Max Crowley	Duke University
22	Prof. Riccardo De Lisa	Università degli studi di Cagliari
23	Alfonso Del Giudice	Università Cattolica del Sacro Cuore
24	Alex R. Dopp **	University of Arkansas
25	Emma Dowling	Middlesex University
26	Robert L Fischer **	Case Western Reserve University
27	Paul Flatau	University of Western Australia
28	Max French	Northumbria University
29	Rusni Hassan	International Islamic University, Malaysia
30	Debra Hevenstone	University of Applied Sciences, Switzerland
31	Leslie Huckfield **	Glasgow Caledonian University
32	Louise Humpage**	University of Auckland
33	David Jamieson	Northumbria University
34	Meghan Joy	Ryerson University
35	Jasper Kim	Ewha Womans University, South Korea
36	Mike Martin	Newcastle University
37	Nicholas Mays	London School of Hygiene & Tropical Medicine, London, UK

38	Neil McHugh	Glasgow Caledonian University
39	Daitaro Misawa	Tokyo Institute of Technology
40	Ali Mollinger-Sahba **	University of Western Australia
41	Daniel Neyland	University of London
42	Sheela Pandey	Pennsylvania State University
43	Francesco Rania	Universita' degli Studi "Magna Græcia" di Catanzaro
44	Francisca García-Cobián Richter **	Case Western Reserve University
45	Daniel Schepis	University of Western Australia
46	Christophe Schinckus	University of the Fraser Valley
47	John Shields	Ryerson University
48	Stephen Sinclair	Glasgow Caledonian University
49	Stefanie Tan	London School of Hygiene and Tropical Medicine, London, UK
50	Jakob Zinsstag **	University of Basel
51	Ian Dewae	Hogent University of Applied Sciences Ghent
52	Dan Edmiston	University of Leeds
53	Celeste Brubaker **	Village Enterprise
54	Suzanne Kerns **	University of Colorado Denver - Anschutz Medical Campus

*** is used to highlight individuals who sent additional articles*

Sample email

Dear _____,

I hope this email finds you well. By way of introduction, I am a PhD student, undertaking the DPhil in Social Intervention & Policy Evaluation at the University of Oxford. As part of my research, I am conducting two systematic reviews (an effectiveness review and a qualitative review) aimed at better understanding how and why Social Impact Bonds have spread throughout various public sectors since 2010.

For the quantitative review, I am collecting impact evaluations of SIB-funded programs that have used either an experimental or quasi experimental design.

For the qualitative review, I am collecting studies, which have: 1) focused on the design and/or implementation of SIBs; 2) used qualitative methods such as interviews, roundtables, focus groups etc; and 3) discussed the benefits, critiques and/or motivations for pursuing SIBs.

As you have been involved in relevant research, I am emailing to see if you would be willing to send me any relevant studies or reports (whether or not they are published) that you are aware of and think should be included.

Thank you in advance for your assistance.

Kind regards,

Vanessa

Appendix E: Quality appraisal – effectiveness review

I used a modified version of the Hamilton tool, to assess the quality of the included studies in the effectiveness review. The small modifications draw on the modified version used in Thomson et al's (2009) review of housing interventions.

The tool consists of six questions, which relate to: 1) selection bias; 2) study design; 3) confounders; 4) blinding; 5) data collection; and 6) withdrawals and dropouts. As described by Lorenc et al (2014), each of the questions can receive either an A (high), a B (medium) or a C (low) quality rating (see table below).

The overall rating for each study is calculated as follows: A = A for Q2 and A/B on at least two of Q1, Q3 and Q6; B = A for Q2 and A/B on one of Q1, Q3 or Q6; or B for Q2 and A/B on at least two of Q1, Q3 and Q6; C = A for Q2 and C for all of Q1, Q3 and Q6; or B for Q2 and A/B on less than two of Q1, Q3 and Q6; or C for Q2.

	Questions	**Rating**
Selection bias	Selected study sample very likely to represent population from target area AND 80–100% response at baseline	A
	Selected study sample very likely to represent population from target area AND 60–79% response at baseline; OR selected study sample somewhat likely to represent population from target area AND 80–100% response at baseline	B
	< 60% baseline response; OR somewhat likely to represent population AND < 80% response; OR not likely to represent population OR representativeness not reported/unclear; OR response rate at baseline not reported/unclear	C
Study design	Control group and pre and post longitudinal data OR random allocation	A
	No control group and pre and post longitudinal data; OR control group and pre and post cross-sectional data AND no indication of major change in population	B
	No control group and pre and post cross-sectional data; OR control group and pre and post cross-sectional data AND possibility of major change in population	C
Confounders	Control group matched on key variables [at least two of crime rate (area level), SES or relevant proxies (area or individual level), gender, age, ethnicity (individual level)] AND supporting data presented; OR outcomes adjusted for key variables (at least two of gender, age, ethnicity, SES) using appropriate methods	A
	Stated that control group matched or 'similar' but supporting data not presented	B

	No matching or adjustment reported AND likely to be substantial differences between groups; OR no information on differences between intervention and control group; OR no control group	C
Blinding	Both outcome assessors AND participants blind to allocation	A
	Either outcome assessors OR participants blind to allocation	B
	Blinding not reported; OR no control group	C
Data collection	Piloting or pretesting of tool; OR checks on validity of data (e.g. verification of a percentage of responses); OR tool shown to be reliable in relevant population	A
	Data collection tool based on previous research but no piloting or checking and reliability not demonstrated	B
	Data collection unclear; OR tools not piloted, checked or based on previous research	C
Withdrawals and dropouts	Attrition < 20%	A
	Attrition 21–40%	B
	Attrition > 40%; OR attrition not reported; OR cross-sectional data only	C

For ‘study design’ question: ‘Longitudinal’ = same individuals pre and post; ‘cross-sectional’ = different individuals. When studies use mixed designs (e.g. presenting both cross-sectional and longitudinal data), give the highest grade applicable to the analyses actually reported. When studies collect longitudinal data and report attrition rates, grade as longitudinal even if only cross-sectional analyses are reported.

For ‘confounders’ question: RCTs will be graded ‘B’ if no information on between-group differences is presented.

For withdrawals and dropouts question: Attrition is measured as the percentage of the baseline sample lost at final follow-up.

Appendix F: Data Extraction Form – Qualitative Systematic Review

I adapted an existing data extraction tool for qualitative research (NICE – data extraction form for qualitative studies). The tool was modified slightly to ensure that sufficient data was extracted on the barriers to and facilitators of rigorous evaluation.

Bibliographic details • Study ID (created by review author) • Review author ID (created by review author) • Citation and contact details • Publication type
Eligibility (if ineligible, note the primary reason for exclusion) Is the study either primary research (any study design) or a systematic review? YES /NO Is the study about the design or implementation of at least one SIB? YES / NO Does the study include a sample of practitioners with relevant SIBs experience? YES / NO Does the study discuss practitioner attitudes regarding motivations for pursuing SIBs? YES / NO Is the study an English document? YES / NO Is the study retrievable? YES / NO
Study aims What was the aim of the study? Is the study linked with any other studies?
Key findings What are the key findings?
Evaluative summary *Draw together brief comments on the study and its strengths and weaknesses. Is further work required? What are its implications for policy, practice and theory?*
Ethical standards Was ethical approval obtained?

Was informed consent obtained? Does the study address ethical issues adequately? Has confidentiality been maintained?
Setting What is the setting for the study? What is the rationale and appropriateness for this choice? Is there sufficient detail about the setting? Over what time period did the data collection take place?
Sample Who was included in the study? Who was excluded from the study? How was the sample selected? Were there any factors that influenced how the sample was selected? What is the size of the sample and groups comprising the study? Is the sample appropriate for the aims of the study, the depth of data it enables to be collected and its breadth?
Data collection What data collection methods were used? Was the data collection adequately described and rigorously conducted? What is the role of the researcher within the setting? Are there any potential conflicts of interest? Is the process of fieldwork adequately described? (if relevant?) How are the data analysed? How adequate is the description of the data analysis? Is adequate evidence provided to support the analysis? Are the researchers' own position, assumptions and possible biases outlined? Are the findings substantiated by the data and has consideration been given to any limitations of the methods or data that may have affected results?
Findings: motivations for uptake of SIBs What instrumental motivations were discussed and/or implied?

What normative (legitimacy) motives were discussed and/or implied? What pragmatic (legitimacy) motives were discussed and/or implied? What regulatory (legitimacy) motives were discussed and/or implied? What cognitive (legitimacy) motives were discussed and/or implied?
Policy and practice What is the country of the study? What is the policy area of focus? To what extent are findings generalisable? What are the implications for policy? What are the implications for practice?
Miscellaneous Funding source? Miscellaneous comments from study authors? Miscellaneous comments by review author? References to any other relevant studies?
Decision Should this study be in the final review? YES NO UNCLEAR

Appendix G: Characteristics of included studies – Effectiveness Review

Study (Publications)	Design / Quality rating / location / policy area	Evaluation objectives	Intervention	Experimental groups	Outcomes measured for SIB payments	Effect size [95% CI]
Riker's Island SIB – 1 x publication (Parsons et al., 2016)	Propensity score matching / B / USA / Recidivism	Determine whether program led to reductions in recidivism for eligible 16- to 18-year-olds admitted to Rikers prison in 2013.	MRT, an evidence-based intervention that focuses on improved decision making. Participants progress through a series of steps with the help of a trained facilitator.	16 - 18-year olds who were eligible to participate in the ABLE program during 2013, compared with matched historical cohort who passed through the jail before the program started (from 2006 to 2010).	Recidivism bed days (RBDs) – i.e (days held in jail in 12 months post release	No effect size provided; not possible to calculate as no SD provided. Non-significant increase (5.1%) for recidivism bed days, between 2006 to 2010 and 2013 (adjusted program effect).
Educate Girls DIB – 1 x publication (ID Insight, 2018)	Cluster RCT / A / India / Education	Estimate the learning and enrolment gains attributable to the intervention, through a clustered randomized control trial.	Comprehensive community intervention to enrol girls into school. Multiple interventions to improve school attendance and prevent drop-outs. Training of volunteers to deliver a child-centric curriculum one to five times a week to boys and girls in Grades 3-5.	332 schools across 282 villages in rural Rajasthan. Pairwise matching of villages based on block, number of eligible schools and similarity.	Learning gains of students enrolled in grades 3-5 and enrolment of out-of-school girls.	Effect size of 0.31 (0.27, 0.35) for learning gains (Appendix 5). P value <0.01. Confidence intervals (CIs) not provided. No effect size for enrolments outcome as the Working Group decided against estimating enrolments in control villages.
Peterborough SIB – 3 x publications (Jolliffe & Hedderman, 2014; Ministry of Justice, 2014; Anders & Dorsett, 2017)	Propensity score matching / B / UK / recidivism	***Jolliffe and Hedderman 2014:*** validate Propensity Score Matching methodology tested in March 2012 (Cave et al., 2012); and determine whether any payment is due to investors for the first cohort**.** ***Ministry of Justice 2014:*** present the final outcomes for cohorts 1 of the PbR pilots for the SIB at Peterborough and the PbR pilot at HMP Doncaster. ***Anders and Dorsett 2017:*** assess whether the matching approach taken by Jolliffe and Hedderman (2014) worked well when applied to Cohort 2 data.	Recidivism program ('The One Service') – aimed at reducing the reoffending among prisoners discharged after serving a sentence of less than 12 months. Contact is made with prisoners pre-release and continued in the community. The intervention is delivered by a mix of paid case workers and volunteers. The combination of activities for each offender is determined by caseworkers, based on individual need.	Up to 10 prisoners from comparator prisons were matched to each Peterborough prisoner. Matches were selected on the basis of the propensity score (which represents probability of being a prisoner at Peterborough rather than elsewhere).	Investors are paid according to how successful the One Service is in reducing reconvictions.	**Jolliffe and Hedderman (2014):** effect size of -0.06 (- 0.12, 0.01) for reconviction events per person. **Ministry of Justice 2014:** Peterborough SIB effect size (CI) as reported above. Doncaster PbR pilot effect size: no effect size provided; not possible to calculate as no SD provided. 12-month re-conviction rate of 52.2% for offenders released from Doncaster during the cohort 1 period (between October 2011 and June 2012) - compares to 58.0% for offenders released in the baseline year (calendar year 2009). **Anders and Dorsett 2017: effect** size of -0.02 (-0.09, 0.05) for reconviction events.
Chicago Child-Parent Center PFS Initiative – 3 x publications (Gaylor et al., 2017; Gaylor et al., 2018; Gaylor et al., 2019)	Propensity score matching / B / USA / Education	To describe the kindergarten readiness outcomes of the first cohort of children in the SIB-CPC project. Ultimately, the evaluation has been designed to measure the impact of the preschool components on children's short- and long-term outcomes.	The CPC model provides full- or part-time high-quality preschool experiences for three- and four-year old children, and combines those educational experiences with family support services and parent engagement activities.	No comparison group for kindergarten readiness and 3rd grade literacy outcomes (payments calculated using outcomes relative to national standards). For special education outcomes, children matched to children with similar characteristics who did not attend preschool in CPS.	Rate of kindergarten readiness; third grade literacy rates; rate at which students are identified with special education needs and placed in special education services (starting in kindergarten).	No effect size available and it could not be calculated due to insufficient detail in the reporting. Broadly, the reports noted the following: **Gaylor et al., 2017**: the special education placement rate in kindergarten for the Cohort 1 intervention group was slightly lower than the rate for the comparison group (4.38% versus 4.94%). This is a difference of .56% in the rate of special education use between the two groups and is an 11% decrease for the treatment group relative to the comparison group. **Gaylor et al., 2018**: the special education rate was lower in the intervention group than the comparison group (3.36% for Cohort 2 in kindergarten and 5.09% for children in the comparison group in kindergarten). This is a difference of 1.73% percent.

						Gaylor et al., 2019: the special education rate was lower in the intervention group than the comparison group (3.95% for Cohort 3 in kindergarten and 6.26% for children in the comparison group in kindergarten). This is a difference of 2.31 percent.
Denver Supportive Housing SIB – 1 x publication (Cunningham et al., 2021)	RCT / C / USA / homelessness	To report housing stability outcomes, in order to understand the impact of supportive housing on jail stays. This brief (and subsequent reports) will be used to determine if the program meets outcome benchmarks as specified in the SIB contract (to know if the city needs to make success payments to the investors).	Supportive housing program for 250 of the city's most frequent users of the criminal justice system. 210 new housing units constructed for use in the SIB; and 40 scattered-site housing units used to house additional people. The program is based upon a proven model that combines the approaches of Housing First with a modified Assertive Community Treatment (ACT) model of intensive case management.	Not discussed in report – apart from noting that individuals were randomized into treatment arm by July 1, 2017.	Payments made based on housing stability outcomes. Factors affecting project's housing stability success payment, were measured (e.g progress toward leasing targets, housing retention and exits, jail stays, progress toward housing stability goals).	No effect size available and could not be calculated as no data were provided on the control group.
London Homelessness SIB – 1 x publication (DCLG, 2017)	Propensity score matching / B / UK / homelessness	Assess whether the intervention (rather than mode of finance) has been successful in: 1) reducing rough sleeping, 2) increasing reconnections and sustainment of those reconnections12, and 3) enabling people to access long-term accommodation over and above what would have happened anyway.	Housing intervention designed around a Navigator approach. Service providers adopted a personalised and flexible approach, supporting the cohort to access existing provision and achieve sustained long-term housing outcomes.	Matched comparison group were selected using the same criteria and time frame but in relation to different years. SIB cohort sample size of 787. Comparison group cohort size of 505.	Reducing rough sleeping, achieving long-term sustained accommodation outcomes, achieving sustained reconnections where appropriate, improving employability, and employment and health outcomes.	No effect size available for reduced rough sleeping or other outcomes; could not be calculated as no SD given. After two years, the mean number of rough sleeping contacts for the SIB group was 9.2 compared to 13.9 for the comparison group (p value 0.0145). 60% of the SIB group slept rough once or more in the two years after the start compared to 67% of the comparison group (p value 0.0407).
Newpin SIB – 2 x publications (Wallace, Hutley, Tomiczek & Batchen 2016; Wallace, Wise & Griffiths 2018)	Live matched control group / B / Australia/ out-of-home-care	None of the reports included an assessment of the SIB financing arrangement. Instead, each report focused on factors such as: program participation, completion and restoration/preservation outcomes; the implementation and ongoing development of the Newpin model etc. Actual counterfactual for control group only available June 2016 onwards.	Intensive child protection and parent education program that aims to break the cycle of destructive family behaviour and to enhance the relationships between parents and children. Program involves: parenting modules; therapeutic group meetings; child development activities and a supportive environment.	Live matched control group was established for Cohort 1 families only. Control group families met the Cohort 1 definition but did not receive Newpin intervention. They received 'business as usual' support, including restoration and other supports.	Investor returns are based on the proportion of children participating in the Newpin program that are restored to their families.	No effect sizes provided; not possible to calculate for any of the reports, as SD's not given. Most recent restoration info available: Between 1 July 2013 and 31 December 2017, 53% of children in the Intervention Group were successfully restored to their families, compared to 18% of children in the Control Group. Testing for statistical significance identified that the net restoration rate for the Intervention Group was significantly higher compared to the Control Group (p <.01).
Resilient Families SIB – 4 x publications (Leahy et al., 2015; Leahy et al., 2016; Leahy et al., 2017 ; Leahy et al., 2020)	Matched control group / B / Australia / out-of-home-care	As above, none of the reports included an assessment of the SIB financing arrangement. Instead, each report aimed to: assess the implementation, effectiveness and cost-effectiveness of the Resilient Families service and to consider the appropriateness of the child protection outcomes used for bond payment purposes.	The RF program is a family support intervention designed for families in which there are concerns about the safety and wellbeing of children which, if not addressed, are likely to result in the removal of the children. The intervention provides therapeutic and practical supports aimed at engaging troubled families and building relationships.	Outcomes for children receiving the intervention are compared to outcomes for similar children in Control group families. The groups are closely matched on a one-to-one basis. Control group families meet the criteria for the service and would have been referred to it if they lived in an agreed location.	Outcome payments are measured through the number of: 1) reports to the Helpline; 2) Safety and Risk Assessments (SARAs) commenced by Family and Community Services (FACS); and 3) entries into statutory OOHC.	No effect size provided by Leahy et al (2015). Not possible to calculate as these studies only include baseline data, for treatment group only (compared to normative data). Leahy et al 2016: effect size for number of helpline reports: 0.16 (-0.21, 0.53). Effect size for SARAs commenced: 0.22 (-0.29, 0.74). No effect size given for entries into OOHC (and cannot be calculated as no SD's provided). The Index Group had fewer entries into statutory OOHC (n=15), compared to the Control Group (n=18), a 17% difference. Leahy et al., 2017: effect size for number of helpline reports (0.61 - 0.56)/1.151086 = 0.043437; effect size for SARAs

						commented (0.3 - 0.3)⁄0.751665 = 0; cannot calculate effect sizes for entries into OOHC (no SD's provided). Evaluation broadly noted that Index group members had fewer OOHC entire (27) than control group members (35). Leahy et al., 2020: on p 81, an effect size of 0.80 is reported for OOHC entry. Effect sizes cannot be calculated for helpline reports or SARAs commenced due to missing information (no standard deviations provided). Evaluation notes that the intervention had a limited impact on the number of Helpline reports (after nine months, both groups received a similar and declining number of reports). It also had a limited impact on the number of SARAs commenced (after nine months, a similar and declining number of SARAs were commenced for both groups).
Innovation Fund Pilot – 1 x publication (Salis et al., 2018)	Propensity score matching / B / UK / employment, education	Illustrate the findings from the quantitative evaluation of the IF pilot, which included two main components: an impact assessment (with two separate analyses – one using survey data and one using admin data) and an SROI analysis. The impact assessment using survey data aimed to explore the impact of the program on the likelihood of participants taking part in various activities related to employment, education and training. The impact evaluation based on administrative data explored the causal effect of the IF pilot on three educational outcomes and two attitudinal/behavioural outcomes.	Innovation Fund pilot comprised of ten projects, which were commissioned in two rounds (Round One started in April 2012 while Round Two commenced operation in November 2012).	The comparison group was obtained by implementing the Propensity Score Matching (PSM) methodology, which entailed pairing (or 'matching') each IF participant with the most similar non-participant young person.	Participation rates in the following activities: 1) school or college; 2) paid job; 3) training course/scheme; 4) apprenticeship; 5) traineeship.	No effect sizes available and cannot be calculated. Results from survey data: difference in % of cohort likely to be in: school or college (-9.8***); paid job (-5.0***); training course/scheme (3.5**); apprenticeship (2.5); traineeship (2.0). *** and **: Statistically significant at the one and five per cent level, respectively.
ON TRACC SIB – 1 x publication (Ooi, Poynton & Halloran 2020)	RCT / X / Australia / Recidivism	To assess the effectiveness of the On TRACC program in reducing recidivism, in comparison to existing services provided to high-risk parolees.	The Transition, Reintegration, and Community Connection (On TRACC) program supplemented existing services provided to high-risk offenders serving a parole supervision order. Service providers delivered enhanced post-release support services to eligible individuals; providers were given a broad remit to deliver the services. Three stages of support were delivered, over a 12 month period post release.	Eligible prisoners were referred to the NSW Bureau of Crime Statistics and Research (BOCSAR) to be randomly assigned to either the treatment or comparison group. Those assigned to the comparison group received existing parole supervision services delivered by NSW Community Corrections. Offenders assigned to the treatment group participated in the On TRACC program and also received regular supervision.	The primary outcome measure (linked to investor repayments) was a reduction in the re-incarceration rate of participating parolees in the 12 months after their release. While there was one primary outcome for repayment purposed, the evaluation included the following four measures of recidivism, including: 1) the probability of return to custody; 2) reconviction; 3) reimprisonment; and 4) committing a personal, property, or serious drug offence within 12 months of release from prison.	Effect sizes not provided but could be calculated from the information contained in the evaluation report (p 9): *Return to custody:* 0.806226 *Reconviction:* 1.240347 *Reimprisonment:* 1.162637 *Personal, property or serious drug offence:* 1.189473 Difference not statistically significant.

Just in Reach SIB – 1 x publication (Hunter et al., 2022)	RCT / X / USA / Recidivism	To provide information about the following: 1) program implementation; 2) county service use and associated costs prior to enrolment, during transitional period and during program period; and 3) estimates of the program impacts.	The Just in Reach SIB provided PSH to individuals in the Los Angeles County jail who were deemed suitable for a clinical diversion intervention which removes people from jail through a court intervention. The program involved a long-term housing subsidy and intensive case management services.	The comparison group consisted of individuals who met the program eligibility criteria (i.e. history of jail time, homelessness, diagnosable chronic condition or behavioral health condition) but were not enrolled in the program and did not receive subsidized permanent housing support during the same time period.	1) housing stability (i.e. the client's retention rate after moving into PSH at the six and twelve month mark); 2) jail avoidance (i.e. the number of rearrests during the two year period following each individual's placement into supportive housing).	Effect sizes not provided and it is not possible to calculate them from the information provided on program effects (p 17-26). Broadly, the JIR PFS program was associated with significant reductions in use of county jail, probation, homelessness, and inpatient physical and mental health services and significant increased use of outpatient mental health services relative to a comparison group.
Resolve SIB – 1 x publication (Wise et al., 2022)	Matched control group / X / Australia / Recidivism	The aim of the report is not outlined explicitly although five areas of investigation are provided, including: 1) implementation; 2) outcomes; 3) innovation; 4) cost effectiveness; and 5) unintended consequences. The evaluation reports on each of these areas.	The Resolve SIB provides a mental health service which blends psycho-social and clinical services to support people living with severe and persistent mental health issues. Beneficiaries receive access to tailored, recovery-oriented support options. Each client can access the program for up to two years from the point of enrolment.	Initially program performance was measured relative to a control group, but measurement was changed in 2021 to a comparison with a client's experience during the one year prior to their enrolment (pre/post program participation measure).	Reduction in the consumption of health services. This is measured as the percentage reduction in National Weighted Activity Units (NWAUs). NWAUs capture an individual's total health related consumption, including both the intensity and duration of the services accessed.	Effect sizes not provided and it is not possible to calculate them from the information provided on program effects (as information not provided on both groups, key statistical details such as means, sample sizes, standard deviations missing). Broadly, however, it notes that in 2020, based on the cumulative NWAU reduction, it was identified that Resolve was underperforming relative to the Control Group.
Silver Chain SIB – 1 x publication (Paxon Group 2020)	Matched control group / X / Australia / Health	The evaluation aims to assess the implementation, effectiveness and cost effectiveness of the Community Palliative Care service. It is the first Implementation Evaluation Report and thus reports on the initial implementation of the service and the provision of services to the first cohort of participants (between 1 July 2017 to 30 June 2018). It outlines key achievements and learnings and identifies opportunities to improve service delivery.	The Community Palliative Care service delivers enhanced community-based palliative care 24/7 in the home. It provides clinical care, practical support for daily activities, support for families and carers and bereavement support. The service was designed for (but is not confined to) people in their last three months of life and supports clients to die in the place of their choice.	The counterfactual consists of a group of comparable individuals who are not using the Silver Chain service. They received usual care.	Outcome payments are linked to the reduction of hospital use by patients receiving the service, relative to the comparison group. This was measured using NWAUs.	Effect sizes not provided but could be calculated from the information contained in Table 21 (p 55). Cohen's d = (5.06 - 4.95)／0.322891 = 0.340672. Overall, individuals in the treatment group experienced a small, but not statistically significant reduction in NWAUs compared to the control group.
Junior Code Academy – 1 x publication (Maze et al., 2018)	Matched control group / X / Portugal / Education	To outline positive outcomes achieved, along with learning and improvement points.	The Junior Code Academy program aimed to improve primary education students educational attainments by teaching them computer coding. It consisted of weekly 2-hour programming classes, offered at primary schools and within school hours in the presence of the head teacher, for 56 weeks of training.	Control group of students who did not receive the intervention.	1) Performance in problem solving and logical thinking; 2) Education attainment in Portuguese language; 3) Education attainment in Mathematics.	Effect sizes not provided and it is not possible to calculate them due to missing information (e.g. means and standard deviations for both groups not provided). Broadly, the report notes that regarding students' logical thinking skills, the treatment groups started from a lower point than the control groups and reached a higher position (2.2% higher at the end of the project). Regarding Portuguese language learning outcomes only assessed at the end of the intervention), the aggregate mean of the treatment group was lower than the aggregate mean of the control groups (-4.56%). For math learning performance outcomes, the aggregate mean of the treatment group was higher than the aggregated mean of the control group (4.10%). The outcomes were not statistically significant.

The Cuyahoga Partnering for Family Success Program – 1 x publication (Crampton et al., 2020)	RCT / X / USA/ Child and family welfare	To evaluate program level outcomes at the conclusion of the five year program.	This program employed Critical Time Intervention (CTI) case management, a family trauma assessment, and created a service plan tailored to each family's needs. Post-reunification, Trauma-Adapted Family Connections therapeutic services were offered and/or traditional family reunification services. Participants were prioritised on housing waiting lists and receive faster housing placement than they would have otherwise.	Control group receiving usual services (although it is unclear what usual services entail).	Total days a child spent in temporary out-of-home placement (OHP)	Effect sizes not provided and could not be calculated from the relevant information provided (p 27-29). Broadly, the study found that treatment group children had a lower rate of exiting custody relative to control group children, although this difference was not statistically significant.
Santa Clara County Project Welcome Home – 1 x publication (Raven, Niedzwiecki and Kushel 2020)	RCT / X / USA/ Child and family welfare	To examine whether randomization to PSH versus usual care reduces the use of acute health care and other services among chronically homeless high users of county-funded services.	This intervention involved PSH (i.e. subsidised housing with closely linked, voluntary supportive services). The additional services included a mix of case management, physical and mental health services and substance use treatment services. PSH is offered on a "housing first" basis; this means that clients are not required to be sober or engage in treatment.	Control group who received usual care. More than one-third of the control group received another form of subsidised housing, thus potentially biasing results. Participants randomized to usual care received referrals to shelters and other homeless services, from staff. They remained eligible to receive all services provided for individuals experiencing homelessness in Santa Clara County, including any form of shelter, and temporary or permanent housing, including PSH not designated for Project Welcome Home.	Outcome payments were tied to the number of months of continuous stable tenancy achieved (3 months, 6 months, 9 months, 12 months, 12+ months). Several additional measures were also included in the evaluation, although not tied to payments. These included:	Effect sizes not provided and could noy be calculated from the relevant information (only treatment group included in Table 2). Broadly, the evaluation found that intervention group members had lower rates of psychiatric ED visits IRR 0.62; 95% CI [0.43, 0.91] and shelter days IRR 0.30; 95% CI [0.17, 0.53], and higher rates of ambulatory mental health services use IRR 1.84; 95% CI [1.43, 2.37] compared to control group members. No differences were found in total ED or inpatient use, or jail time.
South Carolina Nurse-Family Partnership Pay for Success Project – 1 x publication (McConnell et al., 2022)	RCT / X / USA/ Health	To determine the effect of an intensive nurse home visiting program on a composite outcome of preterm birth, low birth weight, small for gestational age, or perinatal mortality.	The program is an established model of nurse home visiting, which involves regular visits beginning prenatally and continuing through 2 postnatal years. In this intervention, nurses provide education, assessments, and goal-setting related to prenatal health, child health and development, and maternal life course.	Control group members received usual care in South Carolina. This included access to all other community and medical services. All study participants were provided with a list of available community resources.	The study includes 3 primary outcomes, although only one has been included in this evaluation (as the others have not yet completed measurement). The primary outcome included in this study is a composite indicating whether a participant experienced any of the following adverse birth events: 1) preterm birth; 2) low birth weight; 3) small for gestational age; or 4) perinatal death. The other 2 primary outcomes include: 1) interbirth intervals of less than 21 months; and 2) major injury or concern for	Effect size could not be calculated for the primary outcome due to missing information (e.g. standard deviation for both groups). Broadly, the evaluation report notes that the incidence of the composite adverse birth outcome was 26.9% in the intervention group, compared 26.1% in the control group (adjusted between-group difference, 0.5% [95% CI, −2.1% to 3.1%]). Intervention group outcomes were not significantly better for any of the maternal and newborn health primary or secondary outcomes. This was the case both for the overall sample and the prespecified subgroups.

					abuse or neglect in the child's first 24 months. There are 54 prespecified secondary outcomes.	
Village Enterprise DIB – 1 x publication (Gallardo et al., 2021)	RCT / X / Kenya and Uganda / poverty reduction	To determine the impact of Village Enterprise's program on household consumption and assets.	The Village Enterprise DIB is a micro-enterprise development program. It operates in a similar manner to an ultra-poor graduation model, as it targets the poorest members of the communities they operate in; they are provided with an integrated package of cash transfers and training.	Control group which was not offered to participate in the Village Enterprise program.	1) Present value of the household increase in consumption; 2) present value of the household increase in net assets	Regarding the consumption outcome, the VE program had a positive and statistically significant impact on monthly consumption for treatment group households that were offered the program. On average, treatment households consumed 9.9 USD (or 6.3%) more per month than the control group. Effect sizes were larger in Kenya (15.2 USD or 7.3%) than in Uganda (3.9 USD or 3.6%). The VE program also had a positive and statistically significant impact on household net assets after the end of the intervention. On average, households in the treatment group have USD 40.5 (or 5.8%) more in net assets than those in the control group. Effect sizes were larger in Kenya (60.9 USD or 8.5%) than in Uganda (15.6 USD or 2.3%). No clear patterns were identified in terms of differential program effects for other pre-specified subgroups (household member with a disability, cohort of treatment household, gender of household head).

Appendix H: Risk of Bias in Studies Referred to in effectiveness review

#	Study	Overall rating	1. Selection bias	2. Study design	3. Confounders	4. Blinding	5. Data collection	6. Withdrawals and dropouts
1	Jolliffe & Hedderman, 2014							
2	Ministry of Justice, 2014							
3	Leahy et al., 2015							
4	Leahy et al., 2016							
5	Parsons et al., 2016							
6	Wallace, Hutley, Tomiczek and Batchen, 2016							
7	Anders & Dorsett, 2017							
8	DCLG, 2017							
9	Gaylor et al 2017							
10	Leahy et al., 2017							
11	Gaylor et al 2018							
12	ID Insight, 2018							
13	Maze et al 2018							
14	Salis et al., 2018							
15	Wallace, Wise and Griffiths, 2018							
16	Gaylor et al 2019							
17	Leahy et al., 2020							
18	Ooi, Poynton & Halloran 2020							
19	Paxon Group 2020							
20	Raven, Niedzwiecki and Kushel 2020							
21	Crampton et al 2020							

22	Gallardo et al 2021							
23	Cunningham et al., 2021							
24	Hunter et al 2022							
25	McConnell et al 2022							
26	Wise et al 2022							



Appendix I: Quality Appraisal using CASP tool ("methodological limitations" section of CERQual Assessment)

Overview of form:

General Study Information

Study Title:

Study Author(s):

Year of Publication:

Questions Instructions:

The first two questions are screening questions and can be answered quickly. If the answer to both is "yes", it is worth proceeding with the remaining questions. For questions 1-9, mark your selected answer with an "X". If you select "can't tell" or have any other thoughts on the questions, please provide notes. For question 10, provide several sentences in response to the question.

1. Was there a clear statement of the aims of the research?

   Yes / No / Can't tell
   Notes:

2. Is a qualitative methodology appropriate?
   Yes / No / Can't tell
   Notes:

3. Was the research design appropriate to address the aims of the research?
   Yes / No / Can't tell
   Notes:

4. Was the recruitment strategy appropriate to the aims of the research?
   Yes / No / Can't tell
   Notes:

5. Was the data collected in a way that addresses the research issue?
   Yes / No / Can't tell
   Notes:

6. Has the relationship between the researchers and the participants been adequately addressed?

   Yes / No / Can't tell
   Notes:

7. Have ethical issues been taken into consideration?
   Yes / No / Can't tell
   Notes:

8. Was the data analysis sufficiently rigorous?

   Yes / No / Can't tell
   Notes:

9. Is there a clear statement of findings?

   Yes / No / Can't tell
   Notes:

10. How valuable is the research?

Yes / No / Can't tell
Notes:

Checklist is available at: https://casp-uk.net/images/checklist/documents/CASP-Qualitative-Studies-Checklist/CASP-Qualitative-Checklist-2018_fillable_form.pdf

Scores for studies analysed in qualitative systematic review ("methodological limitations" section of CERQual Assessment)

#	Study	1	2	3	4	5	6	7	8	9	10	Score (/10)	Quality rating
1	Social Finance (2011)	N	Y	?	?	?	N	N	?	N	Y	2	Low
2	Tomkinson (2012)	Y	Y	Y	N	N	N	N	N	Y	?	4	Low
3	Warner (2013)	Y	Y	Y	N	N	N	N	N	Y	Y	5	Medium
4	Rudd et al., (2013)	Y	?	?	?	?	N	?	?	Y	Y	3	Low
5	Cooper, Graham & O'Dwyer (2013)	Y	Y	Y	N	?	N	N	N	N	?	3	Low
6	Ronicle et al., (2014)	Y	Y	Y	?	?	N	N	?	Y	Y	5	Medium
7	Hughes & Scherer (2014)	Y	Y	Y	?	?	N	N	N	Y	?	4	Low
8	Wallace and Roxas (2014)	Y	Y	Y	Y	?	?	Y	?	Y	Y	7	Medium
9	Disley and Rubin (2014)	Y	Y	Y	?	Y	N	N	Y	Y	Y	7	Medium
10	KPMG (2014)	Y	Y	Y	Y	Y	N	N	Y	Y	Y	8	High
11	Galley, McIsaac & Ymeren (2014)	Y	Y	Y	?	?	N	N	N	?	?	3	Low
12	Mason et al., (2014)	Y	Y	?	?	?	?	N	?	Y	Y	4	Low
13	Rotheroe (2014)	Y	Y	Y	?	?	N	?	?	?	Y	4	Low
14	Reeder et al., (2014)	Y	Y	Y	Y	?	N	N	?	Y	Y	6	Medium

15	Ronicle and Stanworth (2015)	N	?	?	?	?	N	N	?	?	Y	1	Low
16	Ecorys & ATQ Consultants (2015)	Y	Y	?	?	?	?	N	?	N	?	2	Low
17	Disley et al., (2015)	Y	Y	Y	Y	?	?	N	?	Y	Y	6	Medium
18	Ormiston et al., (2015)	Y	Y	Y	Y	?	?	N	Y	Y	Y	7	Medium
19	Reeder et al., (2015)	Y	Y	Y	Y	?	N	N	?	Y	Y	6	Medium
20	Roberts & Cameron (2015)	Y	Y	Y	?	?	N	N	?	Y	Y	5	Medium
21	DCLG (2015)	Y	Y	Y	?	?	N	N	Y	Y	Y	6	Medium
22	Tan et al., (2015)	Y	Y	Y	Y	?	N	Y	Y	Y	Y	8	High
23	Sandstrom & Greenberg (2016)	Y	Y	?	?	?	?	Y	?	Y	Y	5	Medium
24	Ecorys UK & ATQ Consultants (2016)	N	?	?	?	?	N	N	?	Y	Y	2	Low
25	Van Es, Houbem & Nijeholt (2016)	Y	Y	Y	Y	?	?	N	?	Y	Y	6	Medium
26	Ronicle, Fox & Stanworth (2016)	Y	Y	?	?	?	?	N	?	Y	Y	4	Low
27	Griffiths, Thomas & Pemberton (2016)	Y	Y	Y	Y	?	?	N	Y	Y	Y	7	Medium
28	James, Kane & Ravenscroft (2016)	Y	Y	Y	?	?	N	N	N	?	?	3	Low
29	Ecorys UK & ATQ Consultant (2016)	N	?	?	?	?	N	N	?	Y	?	1	Low
30	Cooper, Graham & Himick (2016)	Y	Y	Y	Y	?	Y	N	?	Y	Y	7	Medium
31	CRESSI (2016)	Y	Y	Y	?	?	N	N	N	Y	Y	5	Medium
32	Smeets (2017)	Y	Y	Y	?	?	?	N	?	Y	Y	5	Medium
33	Moffatt et al., (2017)	Y	Y	Y	Y	Y	?	Y	Y	Y	Y	9	High
34	Laing et al., (2017)	Y	Y	?	Y	Y	?	N	Y	Y	Y	7	Medium

35	Hawkins & Bieretz (2017)	Y	?	?	?	?	?	N	?	N	Y	2	Low
36	Hawkins et al., (2017)	Y	Y	?	?	?	?	N	?	Y	Y	4	Low
37	Neyland (2017)	Y	?	?	?	?	?	N	?	Y	Y	3	Low
38	Walsh et al., (2017)	Y	Y	Y	?	?	?	Y	?	N	Y	5	Medium
39	Loxley (2017)	Y	?	?	?	?	N	N	?	Y	Y	3	Low
40	Ronicle et al., (2017)	Y	Y	Y	?	?	?	N	?	Y	Y	5	Medium
41	Edmiston & Nicholls (2018)	Y	Y	Y	Y	?	N	?	?	?	Y	5	Medium
42	Tse & Warner (2018)	Y	Y	?	Y	?	?	N	Y	Y	Y	6	Medium
43	Oroxom, Glassman & McDonald (2018)	Y	Y	?	?	?	?	N	Y	Y	Y	4	Low
44	Fraser et al., (2018)	Y	Y	Y	Y	Y	?	Y	Y	Y	Y	9	High
45	Wainer (2018)	Y	Y	?	?	?	?	N	?	Y	Y	4	Low
46	Williams (2019)	Y	Y	Y	?	?	Y	Y	?	Y	Y	7	Medium
47	ICF Consulting Services (2019)	Y	?	?	Y	?	?	N	?	Y	Y	4	Low
48	Wooldridge, Stanworth & Ronicle (2019)	Y	Y	?	?	?	?	N	?	Y	Y	4	Low
49	Loveless (2020)	Y	Y	Y	Y	?	?	Y	?	Y	Y	7	Medium
50	Rosenback & Carter (2020)	Y	Y	Y	?	?	?	Y	Y	Y	Y	7	Medium
51	Stanworth (2020)	Y	Y	?	?	?	?	N	?	Y	Y	4	Low

Appendix J: Contribution of individual studies on the identified themes

	Study	Instrumental factors	Legitimacy (normative) factors	Legitimacy (pragmatic) factors	Legitimacy (regulatory) factors	Legitimacy (cognitive) factors	Total themes contributed to
1	Social Finance (2011)	Y	Y	Y	N	N	3
2	Tomkinson (2012)	Y	Y	Y	N	Y	4
3	Warner (2013)	Y	Y	Y	Y	Y	5
4	Rudd et al., (2013)	Y	Y	Y	N	Y	4
5	Cooper, Graham & O'Dwyer (2013)	Y	Y	Y	N	Y	4
6	Ronicle et al., (2014)	Y	Y	Y	N	Y	4
7	Hughes & Scherer (2014)	Y	Y	Y	Y	Y	5
8	Wallace and Roxas (2014)	Y	Y	Y	N	Y	4
9	Disley and Rubin (2014)	Y	Y	Y	N	Y	4
10	KPMG (2014)	Y	Y	Y	N	Y	4
11	Galley, McIsaac & Ymeren (2014)	Y	Y	Y	N	Y	4
12	Mason et al., (2014)	Y	Y	Y	N	Y	4
13	Rotheroe (2014)	Y	Y	Y	N	Y	4
14	Reeder et al., (2014)	Y	Y	N	N	Y	3
15	Ronicle and Stanworth (2015)	Y	Y	Y	N	Y	4
16	Ecorys & ATQ Consultants (2015)	Y	Y	Y	N	Y	4
17	Disley et al., (2015)	Y	Y	Y	N	Y	4
18	Ormiston et al., (2015)	Y	Y	Y	N	Y	4
19	Reeder et al., (2015)	Y	N	N	N	Y	2

20	Roberts & Cameron (2015)	Y	Y	Y	N	Y	4
21	DCLG (2015)	Y	Y	Y	N	Y	4
22	Tan et al., (2015)	Y	Y	Y	N	Y	4
23	Sandstrom & Greenberg (2016)	Y	Y	Y	N	Y	4
24	Ecorys UK & ATQ Consultants (2016)	Y	Y	Y	N	Y	4
25	Van Es, Houbem & Nijeholt (2016)	Y	Y	Y	N	Y	4
26	Ronicle, Fox & Stanworth (2016)	Y	Y	Y	N	Y	4
27	Griffiths, Thomas & Pemberton (2016)	Y	Y	Y	N	Y	4
28	James, Kane & Ravenscroft (2016)	Y	Y	N	N	Y	3
29	Ecorys UK & ATQ Consultant (2016)	Y	Y	Y	N	N	3
30	Cooper, Graham & Himick (2016)	Y	Y	Y	N	Y	4
31	CRESSI (2016)	Y	Y	Y	N	Y	4
32	Smeets (2017)	Y	Y	Y	Y	Y	5
33	Moffatt et al., (2017)	Y	Y	N	N	N	2
34	Laing et al., (2017)	Y	Y	Y	N	Y	4
35	Hawkins & Bieretz (2017)	Y	Y	Y	Y	Y	5
36	Hawkins et al., (2017)	Y	Y	Y	N	Y	4
37	Neyland (2017)	Y	Y	Y	N	Y	4
38	Walsh et al., (2017)	Y	Y	Y	N	Y	4
39	Loxley (2017)	Y	Y	Y	N	Y	4
40	Ronicle et al., (2017)	Y	N	Y	N	Y	3

41	Edmiston & Nicholls (2018)	Y	Y	Y	N	Y	4
42	Tse & Warner (2018)	Y	Y	N	N	Y	3
43	Oroxom, Glassman & McDonald (2018)	Y	Y	Y	N	Y	4
44	Fraser et al., (2018)	Y	Y	Y	N	Y	4
45	Wainer (2018)	Y	Y	Y	Y	Y	5
46	Williams (2019)	Y	Y	Y	Y	Y	5
47	ICF Consulting Services (2019)	Y	Y	Y	Y	Y	5
48	Wooldridge, Stanworth & Ronicle (2019)	Y	Y	Y	Y	Y	5
49	Loveless (2020)	Y	Y	Y	N	Y	4
50	Rosenback & Carter (2020)	Y	Y	Y	N	N	3
51	Stanworth (2020)	Y	Y	Y	N	Y	4
	Total (n = number of studies that contributed to final theme)	51	49	46	8	47	Average: 3.94
This table reflects how each study contributed to the construction of the final themes of factors influencing the uptake of SIBs. Note: each study can contribute with multiple factors to the same theme. Shading indicates studies that were judged to be of low quality.							

Appendix K: CERqual Assessment Scores (qualitative systematic review)

	Finding	Adequacy of data[8]	Relevance[9]	Coherence[10]	Methodological limitations[11]	CERQual Assessment
INSTRUMENTAL FACTORS						
1	To improve outcomes	30 / 52.4 / 58.63	Multi (8), multi, yes, 2.53	2.1 density quotes / 2.33 relevance quotes	4.83	Low confidence
2	To achieve greater focus on outcomes	35 / 50.54 / 64.23	Multi (7), multi, yes, 2.31	2.29 density quotes / 2.31 relevance quotes	5.03	High confidence
3	To achieve greater efficiency	14 / 57.36 / 69.3	Multi (7), multi, yes, 2.64	2.07 density quotes / 2.21 relevance quotes	5.21	Moderate confidence
4	To improve service quality	40 / 48.18 / 58.05	Multi (8), multi, yes, 2.23	2.08 density quotes / 2.1 relevance quotes	5.03	Moderate confidence
5	To achieve better value for money	13 / 72.46 / 75.6	Multi (3), multi, no, 2.62	2.35 density quotes / 2.46 relevance quotes	5.08	Moderate confidence
6	To achieve savings	31 / 52.81 / 57.37	Multi (6), multi, yes, 2.58	2.03 density quotes / 2.23 relevance quotes	4.77	Low confidence
7	To achieve better accountability	31 / 59.68 / 75.83)	Multi (6), multi, yes, 2.39	2.19 density quotes / 2.35 relevance quotes	5.26	Moderate confidence
8	To foster better measurement/EBI principles	41 / 48.37 / 61	Multi (6), multi, yes, 2.37	2.10 density quotes / 2.17 relevance quotes	4.88	Low confidence
9	To test the effectiveness of interventions	25 / 55.56 / 67.75	Multi (7), multi, yes, 2.6	2.24 density quotes / 2.48 relevance quotes	4.88	Low confidence

[8] 'Adequacy of data' was calculated by considering the number of reports that contributed to each finding, as well as the size of the reports (page numbers) and average sample size included in the contributing studies.

[9] 'Relevance'was calculated by considering the number of countries and policy areas discussed in the contributing studies, as well as whether contributing studies discussed both developing and developed countries and the relevance of each of the study objectives.

[10] 'Coherence' (i.e. link between data and findings) was calculated by considering the density and relevance of quotes contained in each of the contributing studies.

[11] 'Methodological limitations' was calculated by taking the average CASP score, from each of the studies that contributed to the relevant finding.

10	To test the effectiveness of innovative funding model	9 / 59.44 / 33.2	Multi (4), multi, no, 2.67	2.11 density quotes / 2.22 relevance quotes	4.67	Very low confidence
LEGITIMACY PILLAR ONE: NORMATIVE MOTIVATIONS						
11	To fill a gap in service provision	36 / 49.17 / 64.39	Multi (6), multi, yes, 2.42	2 density quotes, 2.11 relevance quotes	4.56	Low confidence
12	To focus on particularly vulnerable groups	20 / 51.62 / 85.27	Multi (6), multi, yes, 2.35	2.15 density quotes, 2.2 relevance quotes	4.86	Low confidence
13	To shift the focus to early intervention/prevention	6 / 80.67 / 86.5	Multi (4), multi, no, 3	2.5 density quotes, 2.5 relevance quotes	5.12	Low confidence
14	To enhance collaboration	38 / 51.5 / 57	Multi (7), multi, yes, 2.34	2.24 density quotes, 2.26 relevance quotes	5.21	Moderate confidence
15	To expand services	16 / 52.75 / 57.88	Multi (5), multi, no, 2.38	2.25 density quotes, 2.31 relevance quotes	5.25	Moderate confidence
16	To foster innovation	17 / 72.53 / 87.36	Multi (6), multi, yes, 2.71	2.24 density quotes, 2.35 relevance quotes	5.53	High confidence
17	To align with growing/trendy policy area	12 / 64.33 / 91.38	Multi (4), multi, no, 2.5	2.42 density quotes, 2.33 relevance quotes	5.58	Moderate confidence
18	To align with long term governmental reform	11 / 59 / 63.86	Multi (4), multi, no, 2.09	2.36 density quotes, 2.36 relevance quotes	5.45	Moderate confidence
19	To achieve a long-term focus on outcomes	12 / 57.08 / 94.71	Multi (5), multi, no, 2.33	2.42 density quotes, 2.17 relevance quotes	5.83	Moderate confidence
20	To diversify market of providers	3 / 30.67 / 27	Multi (2), multi, no, 2.33	1.67 density quotes, 1.67 relevance quotes	4.67	Very low confidence
21	To stimulate competition	20 / 42.15 / 42.38	Multi (4), multi, no, 2.6	2.15 density quotes, 2.4 relevance quotes	4.8	Low confidence
22	To benefit from private markets/private rigour	30 / 55.93 / 66.56	Multi (6), multi, yes, 2.47	2.1 density quotes, 2.27 relevance quotes	4.73	Low confidence

23	To scale interventions	28 / 45.96 / 53.24	Multi (7), multi, yes, 2.43	2.21 density quotes, 2.25 relevance quotes	4.57	Low confidence
LEGITIMACY PILLAR TWO: PRAGMATIC MOTIVATIONS						
24	To achieve additionality of funding/resources	32 / 52.75 / 66.82	Multi (7), multi, yes, 2.41	2.13 density quotes / 2.25 relevance quotes	4.97	Low confidence
25	To achieve competitive advantage for investors	5 / 74.2 / 99.33	Multi (4), multi, no, 2.2	2.4 density quotes / 2.2 relevance quotes	5.8	Low confidence
26	To allow for the recycling of capital	4 / 111 / 144	Multi (3), multi, no, 3	2 density quotes / 2.25 relevance quotes	4.75	Very low confidence
27	To obtain investor profits	7 / 33.71 / 15	Multi (7), multi, yes, 2.43	1.57 density quotes / 1.57 relevance quotes	5	Low confidence
28	To transfer risk	45 / 49.44 / 59.44	Multi (7), multi, yes, 2.42	2.07 density quotes / 2.22 relevance quotes	4.78	Low confidence
29	To achieve savings for government	31 / 52.81 / 57.37	Multi (6), multi, yes, 2.58	2.03 density quotes / 2.23 relevance quotes	4.77	Low confidence
30	To contribute to CSR/charitable goals/achieve alignment to mission and values	6 / 44.5 / 51	Multi (5), multi, no, 2.33	1.83 density quotes / 1.67 relevance quotes	4.67	Very low confidence
31	To achieve greater funding certainty for NGOs	25 / 56.44 / 82.38	Multi (5), multi, no, 2.52	2.24 density quotes / 2.32 relevance quotes	5.2	Moderate confidence
32	To avoid upfront costs	19 / 59.26 / 50.82	Multi (6), multi, yes, 2.47, 2.52	2.21 density quotes / 2.32 relevance quotes	5.42	High confidence
33	To blend economic and social value	31 / 54.19 / 58.9	Multi (7), multi, yes, 2	2.06 density quotes / 2.03 relevance quotes	4.97	Low confidence
34	To pursue neoliberal agenda/motivations	1 / 20 / unknown	Single, single, no	1 density quotes / 1 relevance quotes	7	Very low confidence
LEGITIMACY PILLAR THREE: REGULATORY MOTIVATIONS						

35	Passing of enabling legislation	5 / 58.6 / 108.5	multi (4), multi, no, 2.4	1.8 density quotes / 1.8 relevance quotes	4.6	Very low confidence
36	Central government support and policy advice	3 / 133.33 / 163.33	multi (3), multi, no, 3	3 density quotes / 3 relevance quotes	5	Low confidence
37	Impact of existing legislation	2 / 140 / 236.5	multi (3), multi, no, 3	3 density quotes / 3 relevance quotes	5.5	Low confidence
LEGITIMACY PILLAR THREE: COGNITIVE MOTIVATIONS						
38	To build capacity and/or build marketplace for impact investment	28 / 59.07 / 64.33	Multi (6), multi, yes, 2.5	2.11 density quotes / 2.21 relevance quotes	5.04	High confidence
39	Evaluation/evidence challenges	39 / 51.28 / 61.83	Multi (6), multi, yes, 2.38	2.18 density quotes / 2.28 relevance quotes	5.18	Moderate confidence
40	Harm/unintended consequences	21 / 54.90 / 58.85	Multi (6), multi, yes, 2.52	2.29 density quotes / 2.38 relevance quotes	5.19	High confidence
41	No clear blueprint/precedent, findings not generalisable etc	25 / 55 / 67	Multi (7), multi, yes, 2.4	1.92 density quotes / 2.04 relevance quotes	5.24	Moderate confidence
42	'SIB effect' unclear	23 / 52.13 / 52.46	Multi (6), multi, yes, 2.52	1.96 density quotes / 2.09 relevance quotes	4.96	Low confidence
43	Lack of appropriate knowledge, skills, resources, funding	27 / 46.93 / 59.69	Multi (7), multi, yes, 2.37	1.81 density quotes / 2.04 quotes releavance	4.89	Low confidence
44	Cost savings unclear, difficult to calculate, focused on too much	24 / 55.75 / 69.6	Multi (6), multi, yes, 2.58	2.21 density quotes / 2.33 relevance quotes	4.79	Low confidence
45	Cautionary academic literature/voices & lack of required literature	13 / 66.69 / 77.13	Multi (5), multi, yes, 2.62	2.15 density quotes / 2.31 relevance quotes	5.31	Moderate confidence
46	Difficulties with testing feasability	12 / 69.42 / 63.71	Multi (5), multi, yes, 2.5	2.12 density quotes / 2.42 relevance quotes	5.5	Moderate confidence
47	Added complexity	35 / 45.17 / 55.47	Multi (6), multi, yes, 2.51	2.11 density quotes / 2.34 relevance quotes	4.6	Low confidence

48	Confusion over aspects of SIBs contracts/PFS	9 / 85.67 / 94.57	Multi (4), multi, yes, 2.89	2.44 density quotes / 2.44 relevance quotes	5.33	Low confidence

Appendix L: Sensitivity analysis (qualitative systematic review)

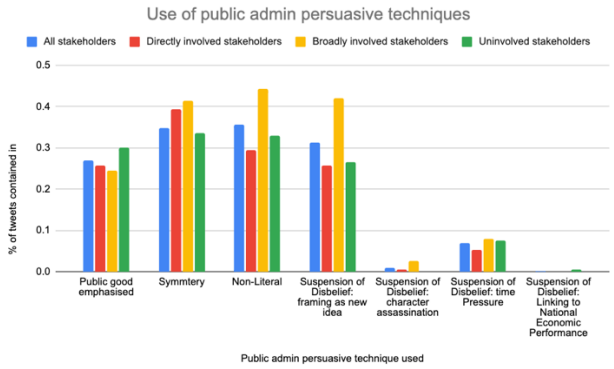
Top reported (instrumental) factors – all studies	Top reported (instrumental) factors – high quality and moderate quality only	Top reported (instrumental) factors – studies published between 2010 - 2013	Top reported (instrumental) factors – studies published between 2014 - 2017	Top reported (instrumental) factors – studies published between 2018 - 2020
To foster better measurement/EBI principles (n = 42)	To foster better measurement/EBI principles (n = 25)	To achieve savings (n = 5)	To foster better measurement/EBI principles (n = 29)	To foster better measurement/EBI principles (n = 10)
To improve service quality (n = 40)	To improve service quality (n = 22)	To improve outcomes (n = 4)	To improve service quality (n = 27)	To improve service quality (n = 10)
To achieve greater focus on outcomes (n = 35)	To achieve better accountability (n = 20)	To foster better measurement/EBI principles (n = 3)	To achieve greater focus on outcomes (n = 25)	To achieve better accountability (n = 9)
To improve outcomes (n = 31)	To achieve greater focus on outcomes (n = 18)	To improve service quality (n = 3)	To improve outcomes (n = 19)	To improve outcomes (n = 8)
To achieve better accountability (n = 30)	To improve outcomes (n = 15)	To test the effectiveness of interventions (n = 3)	To achieve better accountability (n = 19)	To achieve greater focus on outcomes (n = 8)
To achieve savings for government (n = 30)	To achieve savings (n = 15)	N/A	N/A	To achieve savings (n = 8)
N/A	N/A	N/A	N/A	N/A
Top reported legitimacy – (normative) factors – all studies	**Top reported (legitimacy - normative) factors – high quality and moderate quality only**	**Top reported legitimacy (normative) factors – studies published between 2010 - 2013**	**Top reported legitimacy (normative) factors – studies published between 2014 - 2017**	**Top reported legitimacy (normative) factors – studies published between 2018 - 2020**
To enhance collaboration (n = 38)	To enhance collaboration (n = 23)	To fill a gap in service provision (n = 4)	To enhance collaboration (n = 29)	To enhance collaboration (n = 10)
To fill a gap in service provision (n = 35)	To fill a gap in service provision (n = 15)	To stimulate competition (n = 3)	To fill a gap in service provision (n = 22)	To fill a gap in service provision (n = 9)
To benefit from private markets/private rigour (n = 30)	To benefit from private markets/private rigour (n = 14)	To benefit from private markets/private rigour (n = 3)	To benefit from private markets/private rigour (n = 19)	To benefit from private markets/private rigour (n = 9)
To scale interventions (n = 28)	To scale interventions (n = 13)	To scale interventions (n = 3)	To scale interventions (n = 16)	To scale interventions (n = 9)

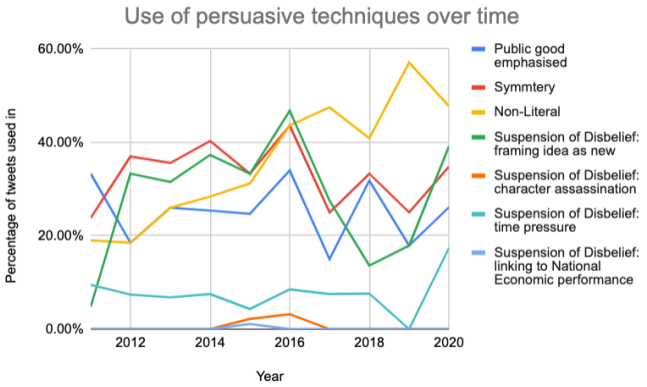
To stimulate competition (n = 20)	To focus on particularly vulnerable groups (n = 11)	To focus on particularly vulnerable groups (n = 2)	To stimulate competition (n = 13)	To foster innovation (n = 7)
To focus on particularly vulnerable groups (n = 20)	N/A	To enhance collaboration (n = 2)	N/A	N/A
N/A	N/A	To expand services (n = 2)	N/A	N/A
Top reported legitimacy – (pragmatic) factors – all studies	**Top reported (legitimacy - pragmatic) factors – high quality and moderate quality only**	**Top reported legitimacy (pragmatic) factors – studies published between 2010 - 2013**	**Top reported legitimacy (pragmatic) factors – studies published between 2014 - 2017**	**Top reported legitimacy (pragmatic) factors – studies published between 2018 - 2020**
To transfer risk (n = 45)	To transfer risk (n = 24)	To achieve savings for government (n = 5)	To transfer risk (n = 29)	To transfer risk (n = 12)
To achieve additionality of funding/resources (n = 32)	To achieve additionality of funding/resources (n = 16)	To transfer risk (n = 5)	To achieve additionality of funding/resources (n = 20)	To achieve additionality of funding/resources (n = 10)
To achieve savings for government (n = 30)	To blend economic and social value (n = 16)	To blend economic and social value (n = 5)	To blend economic and social value (n = 20)	To achieve savings for government (n = 8)
To blend economic and social value (n = 31)	To achieve savings for government (n = 15)	To achieve additionality of funding/resources (n 3)	To achieve savings for government (n = 17)	To achieve greater funding certainty for NGOs (n = 6)
To achieve greater funding certainty for NGOs (n = 25)	To achieve greater funding certainty for NGOs (n = 15)	To achieve greater funding certainty for NGOs (n = 2)	To achieve greater funding certainty for NGOs (n = 17)	To blend economic and social value (n = 6)
N/A	N/A	To avoid upfront costs (n = 2)	N/A	N/A
N/A	N/A	N/A	N/A	N/A
Top reported legitimacy – (regulatory) factors – all studies	**Top reported legitimacy – (regulatory) factors – high quality and moderate quality only**	**Top reported legitimacy (regulatory) factors – studies published between 2010 - 2013**	**Top reported legitimacy (regulatory) factors – studies published between 2014 - 2017**	**Top reported legitimacy (regulatory) factors – studies published between 2018 - 2020**
Passing of enabling legislation (n = 5)	Passing of enabling legislation (n = 3)	Passing of enabling legislation (n = 1)	Passing of enabling legislation (n = 3)	Central government support and policy advice (n = 3)
Central government support and policy advice (n = 3)	Central government support and policy advice (n = 1)	N/A	N/A	Passing of enabling legislation (n = 1)
Impact of existing legislation (n = 2)	Impact of existing legislation (n = 1)	N/A	N/A	Impact of existing legislation (n = 2)
N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A	N/A

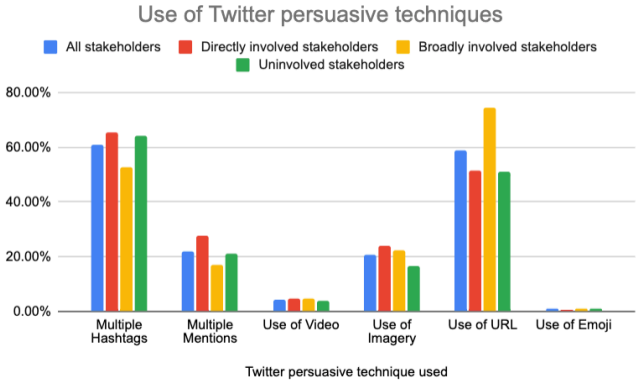
N/A	N/A	N/A	N/A	N/A
Top reported legitimacy – (cognitive) factors – all studies	**Top reported legitimacy – (cognitive) factors – high quality and moderate quality only**	**Top reported legitimacy (cognitive) factors – studies published between 2010 - 2013**	**Top reported legitimacy (cognitive) factors – studies published between 2014 - 2017**	**Top reported legitimacy (cognitive) factors – studies published between 2018 - 2020**
Evaluation/evidence challenges (n = 39)	Evaluation/evidence challenges (n = 24)	Evaluation/evidence challenges (n = 4)	Evaluation/evidence challenges (n = 26)	Evaluation/evidence challenges (n = 10)
Added complexity (n = 35)	Added complexity (n = 19)	Cost savings unclear (n = 4)	Added complexity (n = 25)	To build capacity (n = 8)
To build capacity and/or build marketplace for impact investment (n = 29)	To build capacity and/or build marketplace for impact investment (n = 14)	To build capacity (n = 3)	Lack of appropriate knowledge, skills, resources, funding (n = 21)	Cost savings unclear (n = 8)
Lack of appropriate knowledge, skills, resources, funding (n = 27)	Lack of appropriate knowledge, skills, resources, funding (n = 14)	Harm/unintended consequences (n = 3)	To build capacity (n = 18)	Harm/unintended consequences (n = 7)
No clear blueprint/precedent, findings not generalisable etc (n = 25)	Harm/unintended consequences (n = 12)	Added complexity (n = 3)	No clear blueprint (n = 16)	No clear blueprint (n = 7)
N/A	No clear blueprint/precedent, findings not generalisable etc (n = 12)	N/A	N/A	Added complexity (n = 7)
N/A	Cost savings unclear (n = 12)	N/A	N/A	N/A

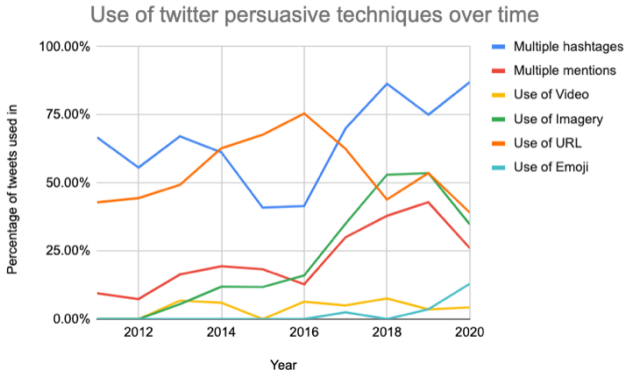
Appendix M: Sensitivity Analysis - Social Media Analysis (Random Sample)

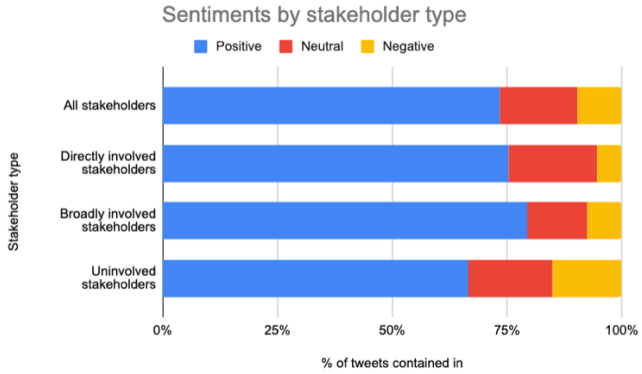
1. Results when recalculated with tweets containing broken links removed

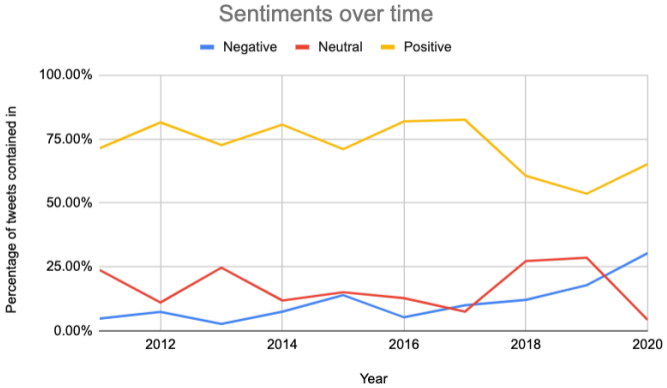


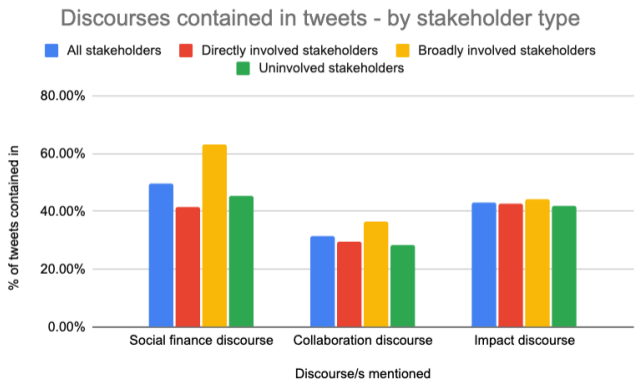


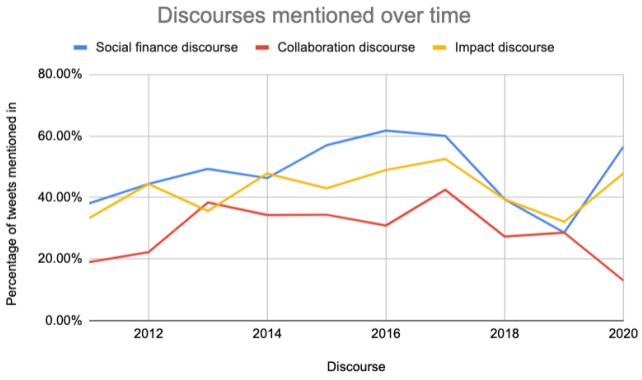


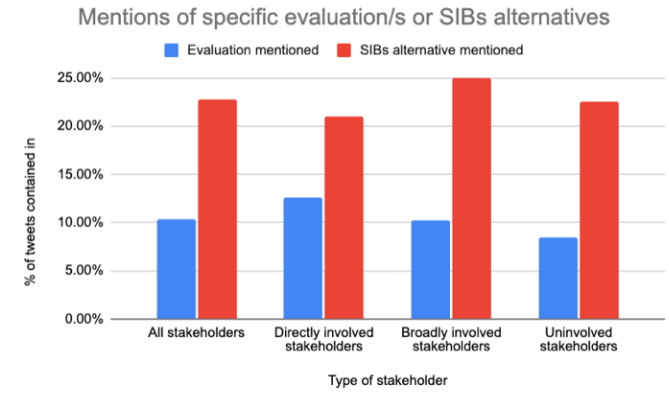


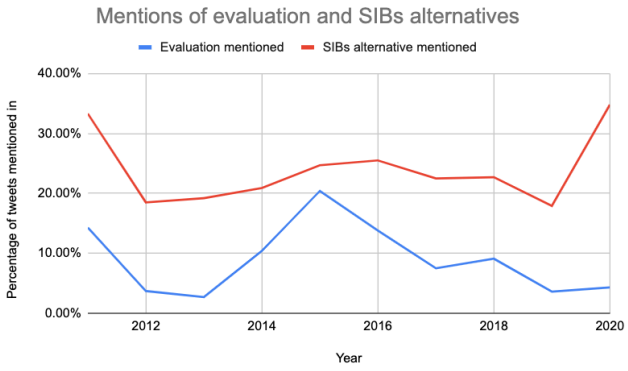




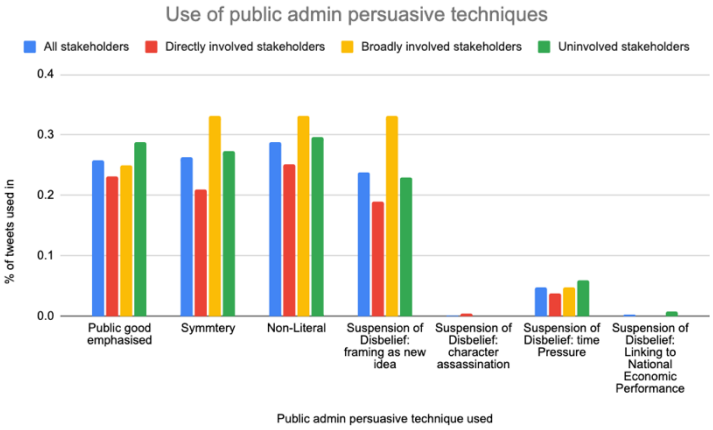


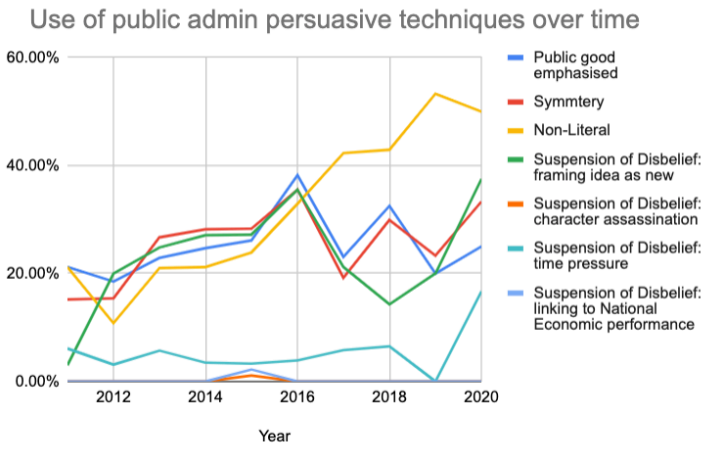


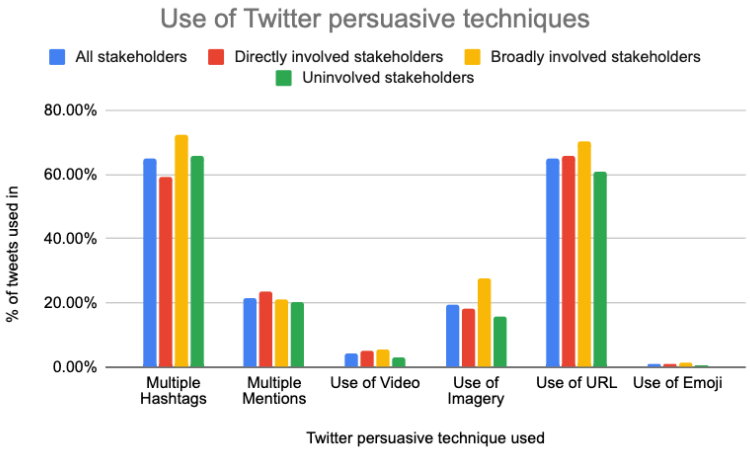


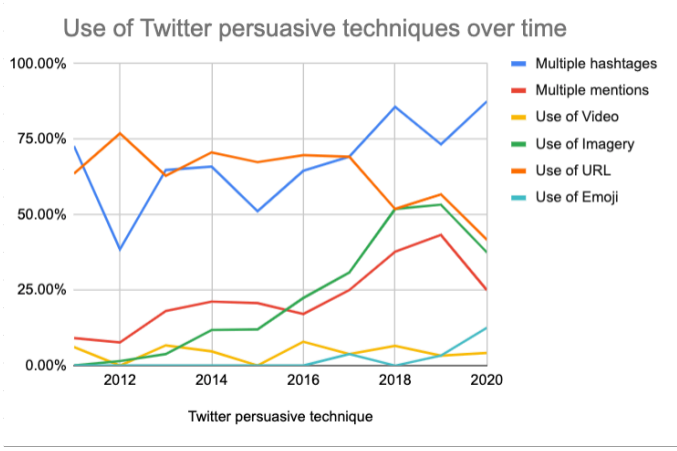


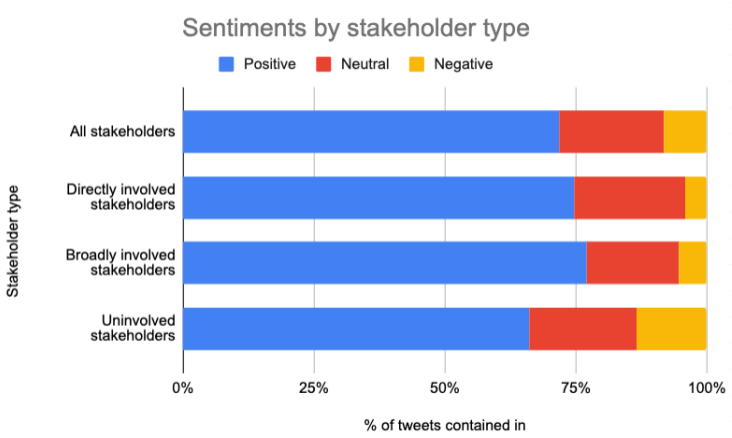
2. *Results when particularly over-represented accounts removed*

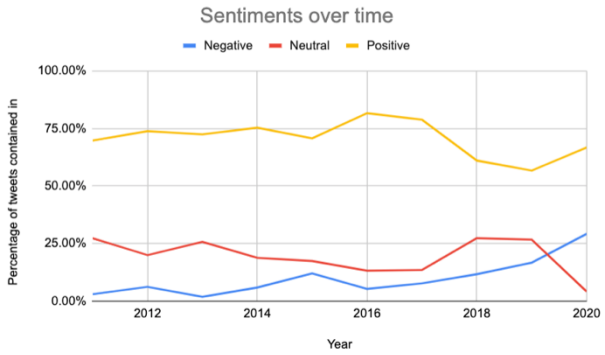


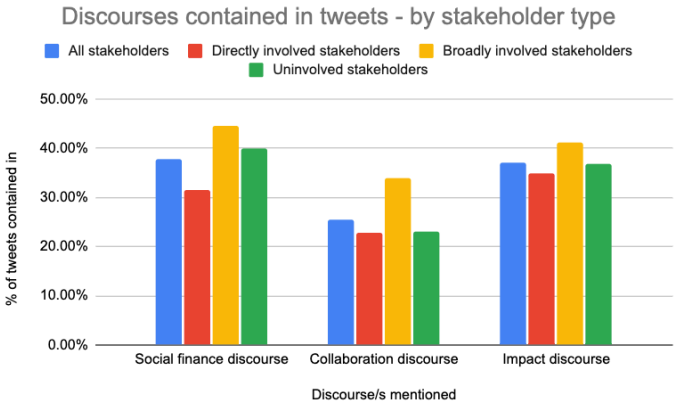


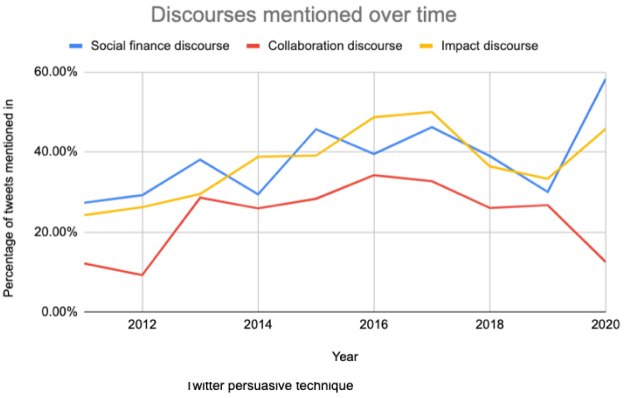

Use of Twitter persuasive techniques
All stakeholders
Directly involved stakeholders
Broadly involved stakeholders
Uninvolved stakeholders
80.00%
60.00%
40.00%
20.00%
0.00%
% of tweets used in
Multiple Hashtags
Multiple Mentions
Use of Video
Use of Imagery
Use of URL
Use of Emoji
Twitter persuasive technique used

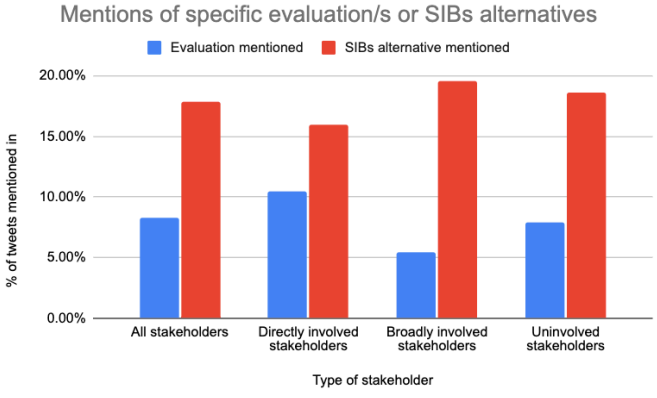

Use of Twitter persuasive techniques over time
100.00%
75.00%
50.00%
25.00%
0.00%
2012
2014
2016
2018
2020
Twitter persuasive technique
Multiple hashtages
Multiple mentions
Use of Video
Use of Imagery
Use of URL
Use of Emoji

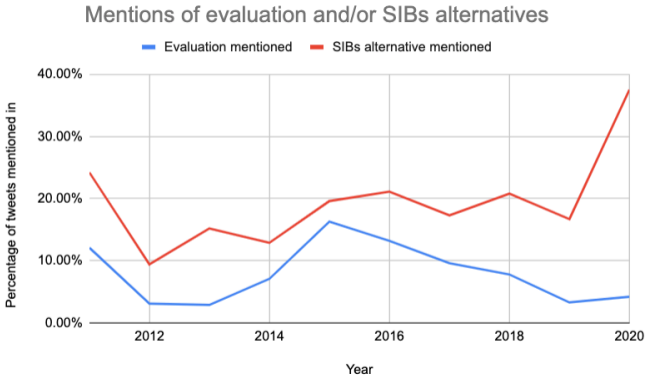

Sentiments by stakeholder type
Positive
Neutral
Negative
All stakeholders
Directly involved stakeholders
Broadly involved stakeholders
Uninvolved stakeholders
Stakeholder type
0%
25%
50%
75%
100%
% of tweets contained in


Sentiments over time
Negative
Neutral
Positive
100.00%
75.00%
50.00%
25.00%
0.00%
Percentage of tweets contained in
2012
2014
2016
2018
2020
Year

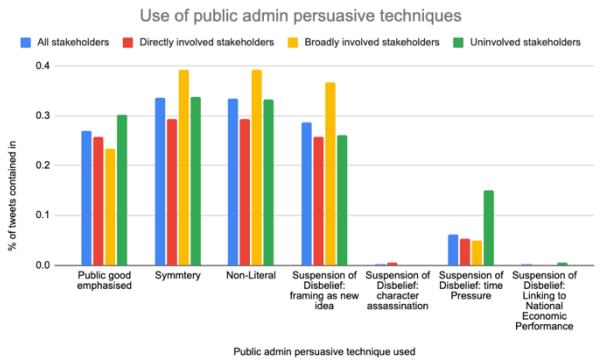

Discourses contained in tweets - by stakeholder type
All stakeholders
Directly involved stakeholders
Broadly involved stakeholders
Uninvolved stakeholders
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
% of tweets contained in
Social finance discourse
Collaboration discourse
Impact discourse
Discourse/s mentioned

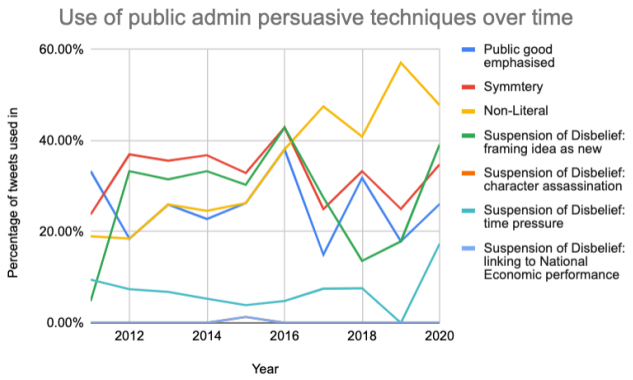

Discourses mentioned over time
Social finance discourse
Collaboration discourse
Impact discourse
60.00%
40.00%
20.00%
0.00%
Percentage of tweets mentioned in
2012
2014
2016
2018
2020
Year
Twitter persuasive technique

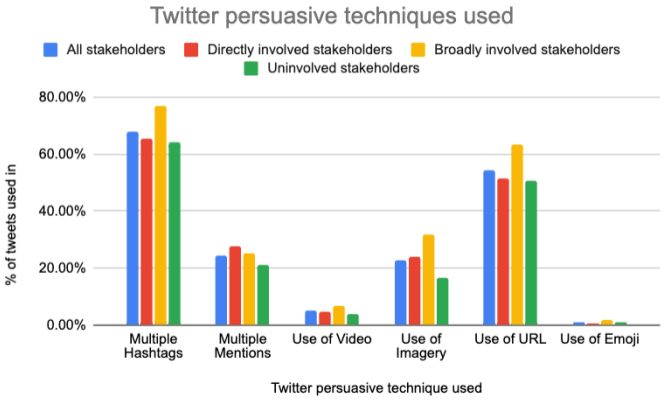

Mentions of specific evaluation/s or SIBs alternatives
Evaluation mentioned
SIBs alternative mentioned
20.00%
15.00%
10.00%
5.00%
0.00%
% of tweets mentioned in
All stakeholders
Directly involved stakeholders
Broadly involved stakeholders
Uninvolved stakeholders
Type of stakeholder

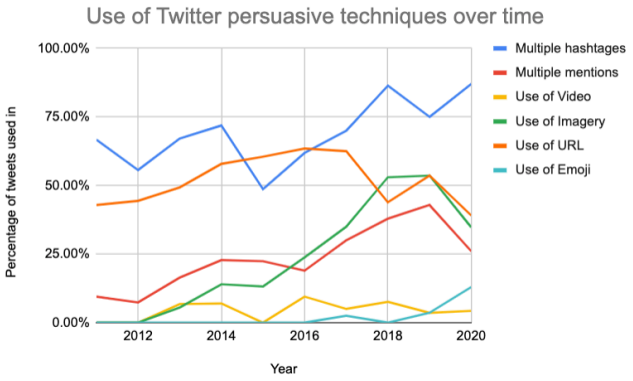

Mentions of evaluation and/or SIBs alternatives
Evaluation mentioned
SIBs alternative mentioned
40.00%
30.00%
20.00%
10.00%
0.00%
Percentage of tweets mentioned in
2012
2014
2016
2018
2020
Year

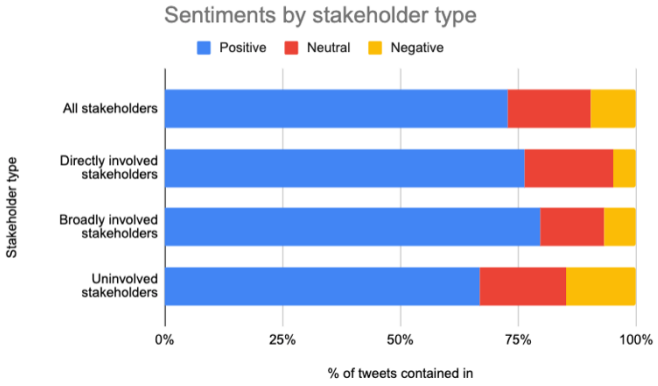
3. *Results when broken links and over-represented accounts removed*

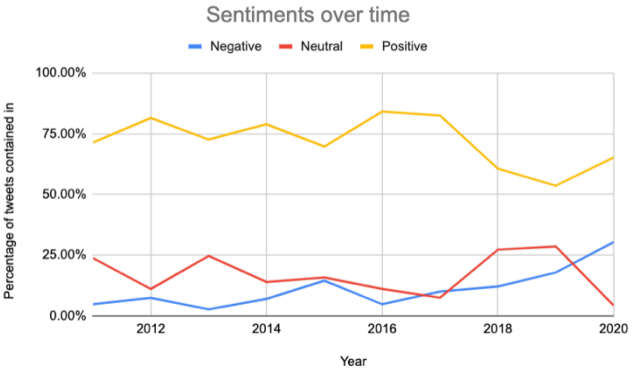


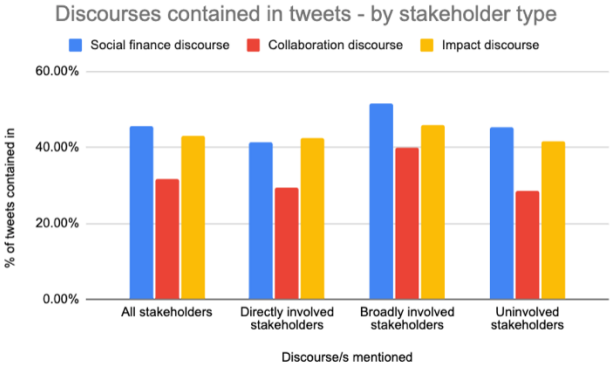


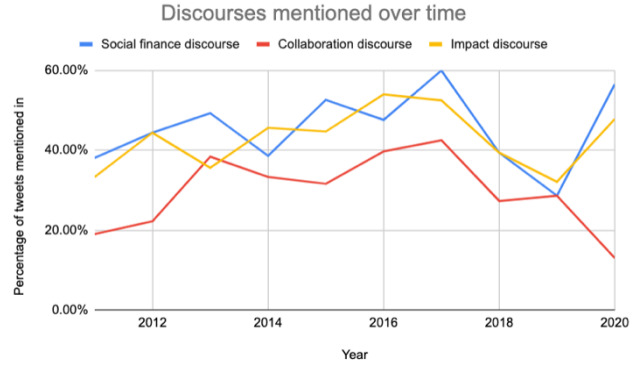


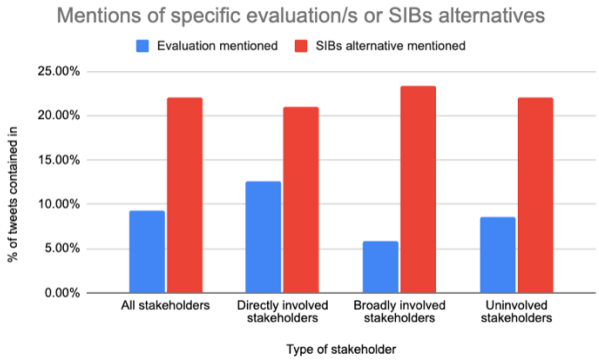


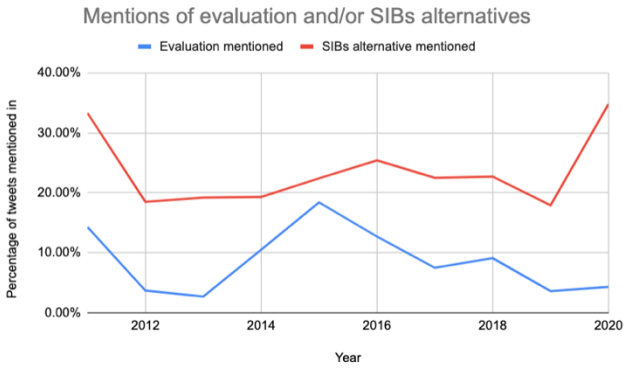




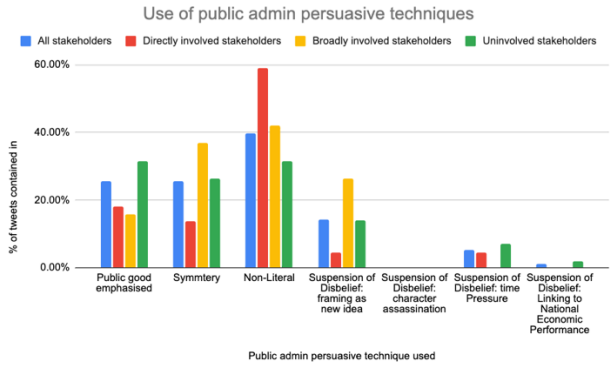


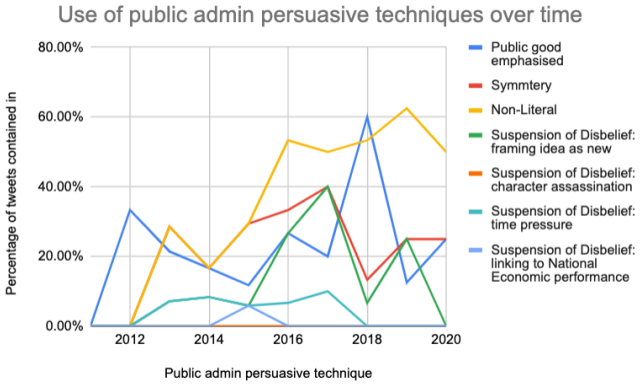


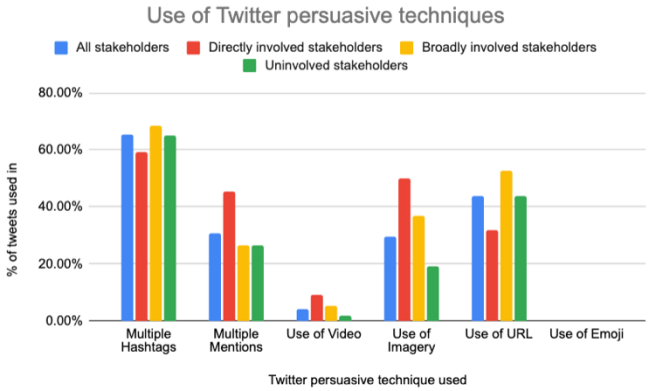


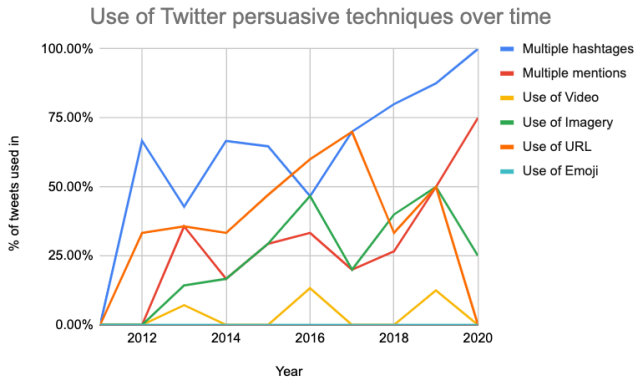


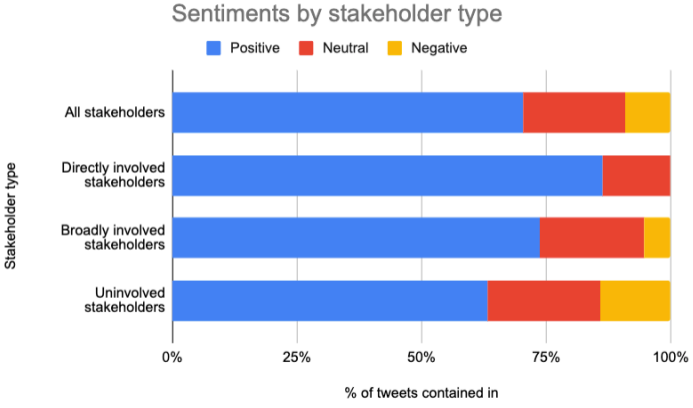
4. *Results when restricted to non-English tweets only*

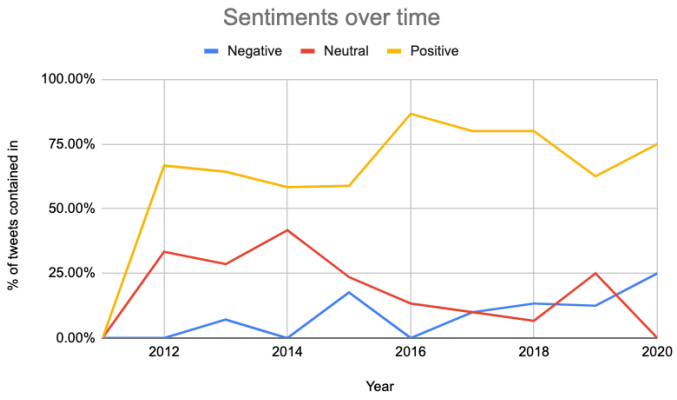


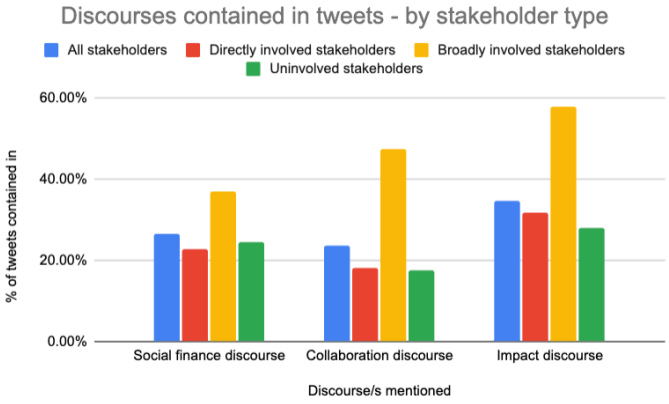


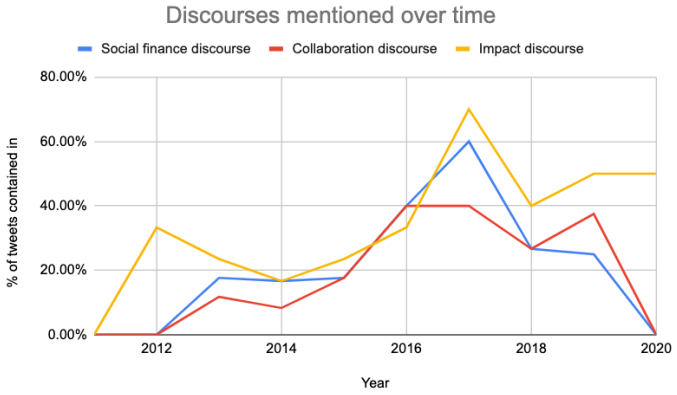


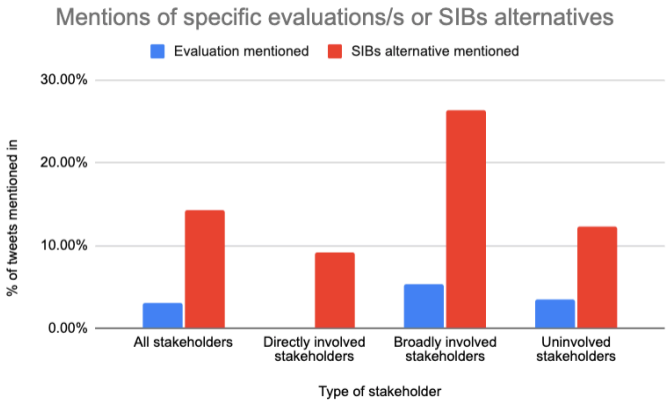


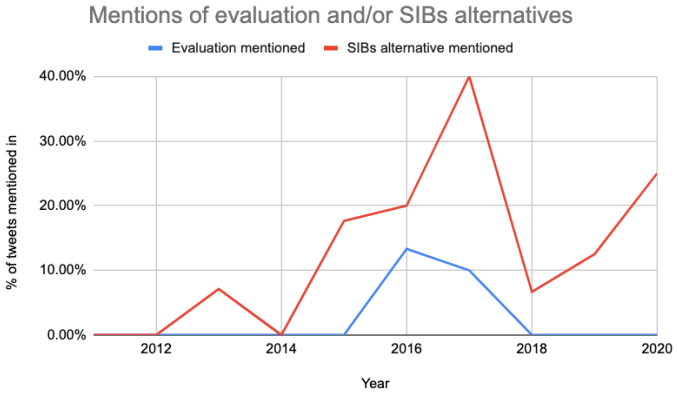




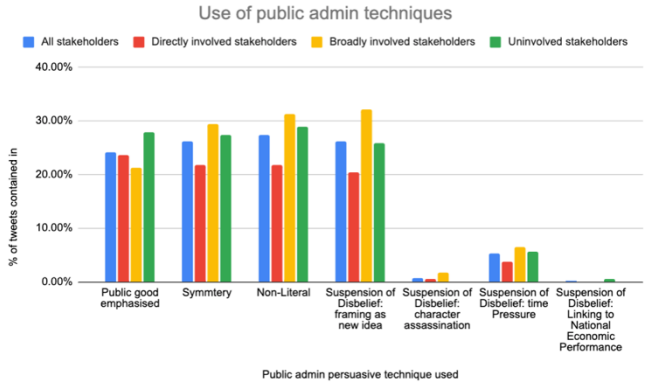


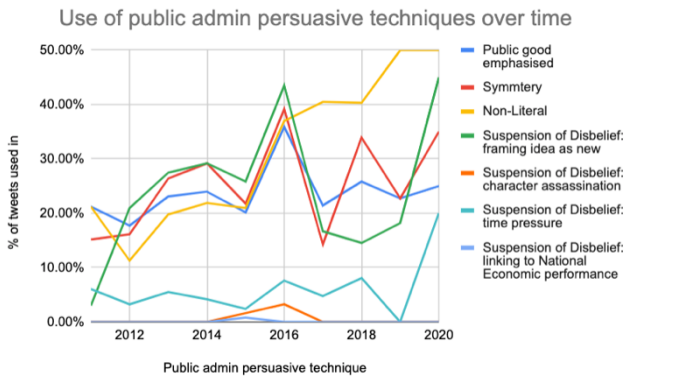


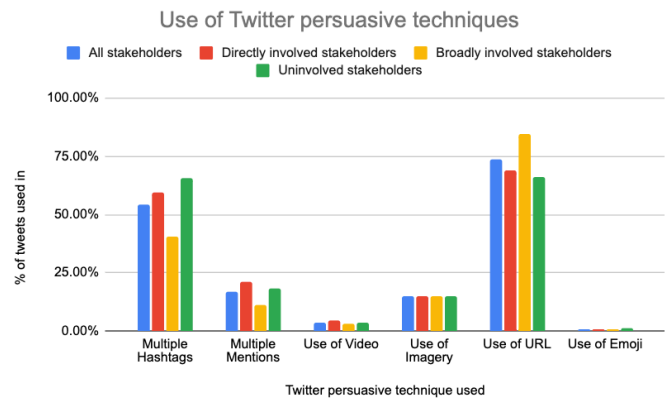


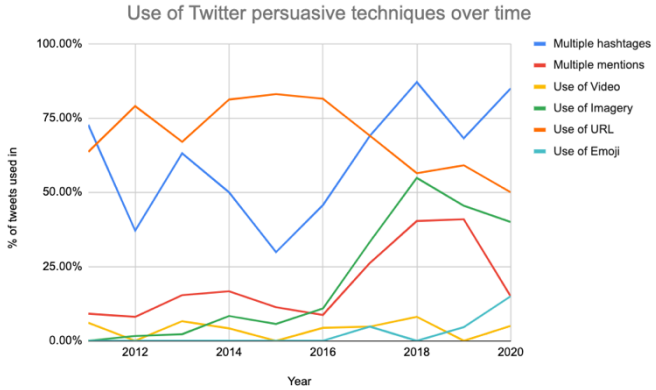


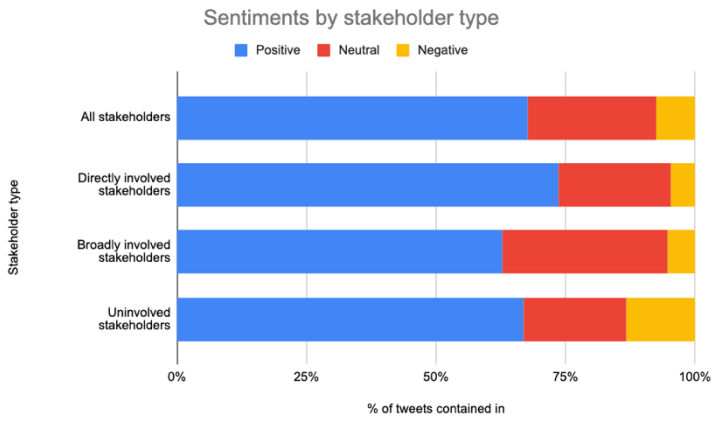
5. Results when restricted to English tweets only

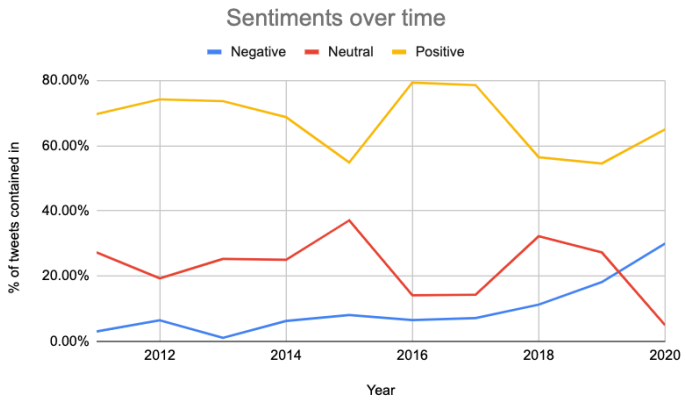


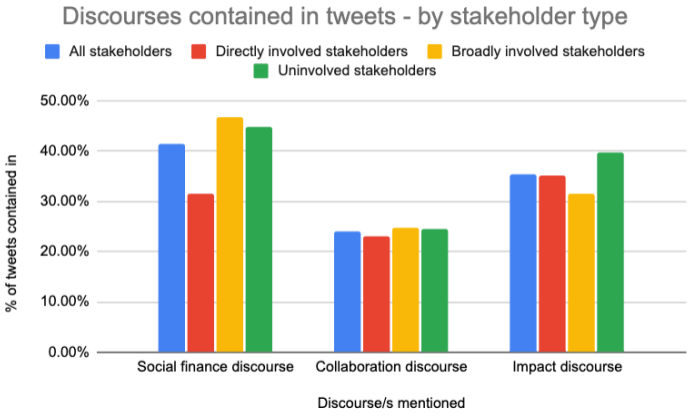


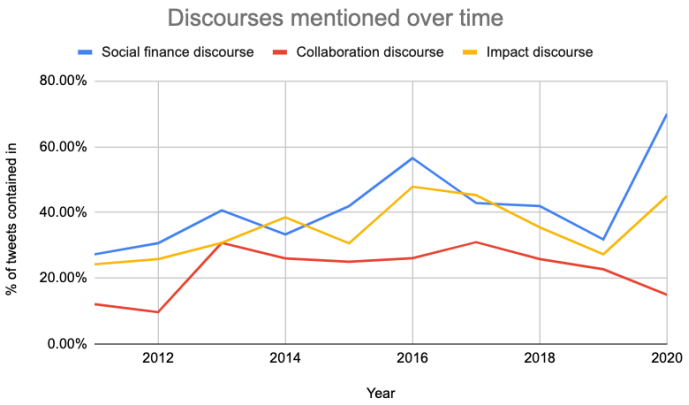


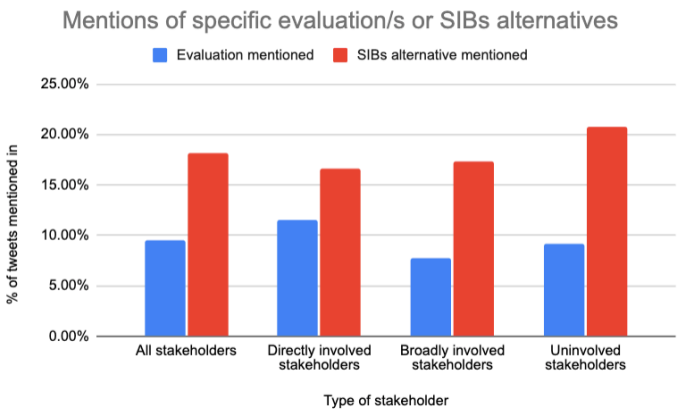


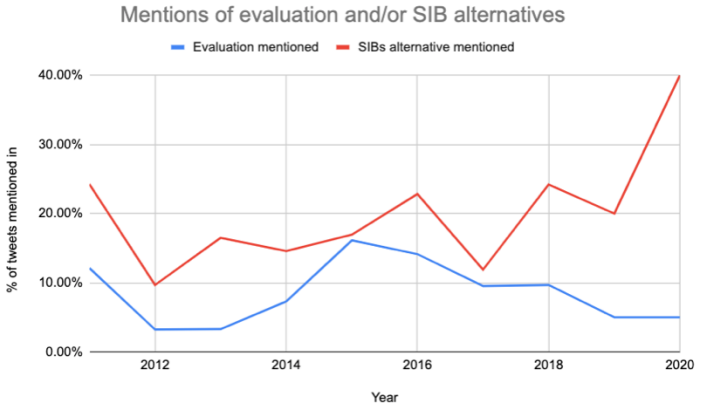

Sentiments by stakeholder type
Positive
Neutral
Negative
All stakeholders
Directly involved stakeholders
Broadly involved stakeholders
Uninvolved stakeholders
Stakeholder type
0%
25%
50%
75%
100%
% of tweets contained in


Sentiments over time
Negative
Neutral
Positive
80.00%
60.00%
40.00%
20.00%
0.00%
% of tweets contained in
2012
2014
2016
2018
2020
Year


Discourses contained in tweets - by stakeholder type
All stakeholders
Directly involved stakeholders
Broadly involved stakeholders
Uninvolved stakeholders
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
% of tweets contained in
Social finance discourse
Collaboration discourse
Impact discourse
Discourse/s mentioned


Discourses mentioned over time
Social finance discourse
Collaboration discourse
Impact discourse
80.00%
60.00%
40.00%
20.00%
0.00%
% of tweets contained in
2012
2014
2016
2018
2020
Year


Mentions of specific evaluation/s or SIBs alternatives
Evaluation mentioned
SIBs alternative mentioned
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
% of tweets mentioned in
All stakeholders
Directly involved stakeholders
Broadly involved stakeholders
Uninvolved stakeholders
Type of stakeholder


Mentions of evaluation and/or SIB alternatives
Evaluation mentioned
SIBs alternative mentioned
40.00%
30.00%
20.00%
10.00%
0.00%
% of tweets mentioned in
2012
2014
2016
2018
2020
Year