Designing the Self-Report of Wellbeing in Pregnancy

A Thesis Submitted to the School of Computer Science & Statistics, Trinity College Dublin for the Degree of Doctor of Philosophy
Supervised by Dr. Gavin Doherty

Kevin Doherty
March 20, 2019
Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I agree to deposit this thesis in the University’s open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Kevin Doherty
March 20, 2019
Acknowledgements

I would like to thank my supervisor, Gavin Doherty, for the time and attention dedicated to the work of this thesis. It would not have been possible without your tenacity and insight.

I must also thank our colleagues José Marcano-Belisario, Martin Cohn, Nikolas Mastellos and Josip Car of Imperial College London and Cecily Morrison of Microsoft Research Cambridge. Working with such a diverse and engaged group of people has been a uniquely fulfilling experience.

Many thanks are due also to the hundreds of women and tens of healthcare professionals across the UK who so willingly gave of their time, experience and expertise.

To my colleagues and fellow students, Erwan Moreau, Akira Hayakawa, Milena Lopes, Christian Saam, Francesca Bonin, Romina Charifou, Carmen Klaussner, Justine Reverdy, Emer Gilmartin, Fahim Salim, Fasih Haider, Maria Koutsombogera, Loredana Cerrato, Alfredo Maldonado, Joao Cabral, Bilal Yousuf, Benjamin Cowan, Camille Nadal, Andreas Balaskas, Zaynab Salman, Alexander Mertens, Nick Campbell and Ashjan Alsulaimani. Thank you for your friendship and support. Special thanks are due also to Marguerite Barry and Bérenger Arnaud for your contributions to this work, your energy and your support. To my friends Claire Gibbons-Cafferkey, Daniel Keenaghan, Raphaelle Giuge and Ciarán Gethings. In many ways, you made this possible.

Finally, thank you to my family, parents Ann and Brian, and brothers Ciarán and Columb who have supported me throughout these years.

I would like to dedicate this work to Derek Murphy.

You lived life in all its colour.

Kevin Doherty

University of Dublin, Trinity College

March 20, 2019
You might say that every day that went by brought each man, provided he did not die, closer to the end of his troubles. Rieux had to acknowledge that this was true, but as truths go, it was rather a broad one.

*Albert Camus | The Plague*
Abstract

At the present moment, one in six adults living in England meets the diagnostic criteria for anxiety or depression. The significant and long-lasting intergenerational effects of maternal mental health in particular make perinatal wellbeing a public health priority. Every year, up to 15% of women are diagnosed with depression during pregnancy or within a year of giving birth. Perinatal depression affects women’s quality of life, birth outcomes and children’s emotional, cognitive, behavioural and social development. Suicide is the leading cause of maternal mortality within the United Kingdom (UK).

Treatment and support needs to be made available to those who require it, but in order to do so, effective programs of assessment, particularly for those at risk of distress, must be in place. In the context of the UK’s National Health Service (NHS), mental health screening in pregnancy is currently carried out verbally and using paper-based questionnaires completed in waiting rooms. However, it is estimated that at least 50% of perinatal depression cases go undiagnosed. Although 96% of midwives report asking women about their mental health during their first appointment, only one in ten women recall being asked.

Mobile devices have the potential to facilitate the self-report and remote screening of wellbeing and depression throughout the antenatal period, extending care to under-served and at-risk groups, enabling timely assessment and intervention, gathering ecologically valid and longitudinal data, overcoming stigma, supporting disclosure, and fostering trust between women and midwives. However, we know little about the feasibility of employing mobile technologies for public health screening in practice, and significant design challenges currently impede their deployment.

This thesis examines the design of mobile applications for the self-report of psychological wellbeing in pregnancy. It asks how technology shapes the self-report of wellbeing, how users engage and are engaged in the honest disclosure of mental health concerns, how health professionals might act upon reports of psychological wellbeing, and how technology might contribute to our evolving conception of wellbeing and its pursuit. Through collaboration with public health researchers, pregnant women and a variety of health professionals including midwives, this thesis seeks to establish a case for the feasibility
of deploying mobile technologies for the self-report of psychological wellbeing and depression during pregnancy.

This thesis contributes knowledge concerning the conception, theory, measurement and design of wellbeing, self-report, user engagement and ecological momentary assessment technologies. It provides insight into the experience of wellbeing and care in pregnancy as voiced by women, practice and research midwives, medical researchers, clinical studies officers, psychologists, psychiatrists, general practitioners, and clinicians in maternal and child health, obstetrics and midwifery. It details the design and evaluation of BrightSelf, a mobile application and online platform for the self-report of wellbeing and depression during pregnancy, and presents the results of the first longitudinal clinical deployment of a mobile technology for antenatal mental health screening.

Using BrightSelf, 355 pregnant women attending 14 NHS midwifery clinics across England provided 2,280 momentary and retrospective reports of their wellbeing in daily life over a 9 month period. Women installed and engaged with this mobile application regardless of their age, education, number of children, marital status, employment status, past diagnosis of depression or level of wellbeing. Thirty-nine women reported a risk of depression, self-harm or suicide using this technology and received immediate midwife support. Two-thirds of participants who received support in this way registered no risk of depression according to the standard screening methods employed in-clinic at baseline, and women spoke positively of the experience;

“Everyone should have access to the app as soon as they find out they're pregnant, great way of communicating, especially for those less inclined to talk to anyone.” Participant No. 349 | Arm 1 | Writing in the Post-Study Survey

This thesis resulted in a new understanding of the feasibility and design of mobile technologies for public health screening and research.
Related Publications

Journal Papers


Conference Papers


Conference Posters


Ethics Submissions


# Contents

Declaration iii  
Acknowledgements v  
Abstract ix  
Related Publications xi  
List of Figures xxi  
List of Tables xxv  
List of Abbreviations xxv

1 Introduction 1  
1.1 Mental Health & Wellbeing ................................................. 1  
1.1.1 Wellbeing ................................................................. 1  
1.1.2 The Public Health Approach to Mental Healthcare .................... 2  
1.1.3 Mental Health in the Perinatal Context ................................ 2  
1.1.4 The Self-Report of Wellbeing ........................................ 3  
1.2 Technology for the Self-Report of Wellbeing ................................. 3  
1.2.1 Engaging Women & Midwives ........................................ 4  
1.3 The Research Question ..................................................... 4  
1.4 Thesis Structure ............................................................ 5  
1.5 An Overarching Methodology | The Epistemology of Design ................. 7  
1.5.1 Designing for Complexity | A Wicked Problem ........................... 7  
1.5.2 Design as Research | Towards What Works ............................... 7  
1.5.3 Critical Realism | An HCI Epistemology ................................. 8
2 Wellbeing & Mental Health

2.1 Defining The Life Well-Lived

2.2 Wellbeing as The Absence of Suffering

2.3 Wellbeing as Functioning

2.3.1 Wellbeing as Wellness

2.4 Wellbeing as Subjective Experience

2.4.1 Wellbeing as Emotion

2.4.2 Wellbeing as Appraisal

2.5 Wellbeing as Flourishing

2.6 Technologies for Wellbeing

2.6.1 Technology, Genes & The Early Environment

2.6.2 Technology & Our Circumstances

2.6.3 Technology, Our Outlook & Our Activities

2.7 Fostering Wellbeing | The Public Health Approach

2.7.1 Pregnancy | An Opportunity to Improve Population Health

2.8 Conclusion | Reflections & Takeaways

3 Understanding Subjective Experience

3.1 The Temporality of Wellbeing

3.2 Articulating Experience

3.3 Ecological Momentary Assessment

3.4 Making Sense of Experience in HCI

3.5 Literature Review Methodology

3.6 The Consolidation of Self

3.6.1 The Experiencing Self

3.6.2 The Remembering Self

3.6.3 The Future-Oriented Self

3.7 The Construal of Experience

3.7.1 Momentary Reflection

3.7.2 Retrospective Reflection

3.7.3 Prospective Reflection
### 3.7.4 Multiple Consciousness

3.8 Conflicting Self-Reports

3.8.1 Ground Truth & Self Truths

3.8.2 Self Knowledge \([d_{OE}, d_{ER}, d_{OR}]\)

3.8.3 Future Choice \([d_{RF}, d_{EF}, d_{OF}]\)

3.8.4 The Drive to Reduce Dissonance

3.9 Designing for Multiple Selves

3.9.1 The Importance of Multiple Time-Frames

3.9.2 Drawing on Self-Knowledge

3.9.3 A Richer Understanding of Experience through Self-Report

3.9.4 Tools for Reflection

3.9.5 Evoking the Selves in Design

3.10 Conclusion | Reflections & Takeaways

### 4 The Design of Self-Report Technologies

4.1 Technology's Reach

4.2 Fulfilling the Promise of EMA

4.3 The EMA Technology Design Space

4.3.1 Modes of Use

4.3.2 EMA Technology Features

4.4 Designing EMA Protocols

4.4.1 Time-Contingent Sampling

4.4.2 Event-Contingent Sampling

4.4.3 Sampling Duration

4.4.4 Questions & Questionnaires

4.4.5 Reactivity, Habituation & Attrition

4.4.6 Validity in Daily Life

4.4.7 Assessment & Intervention

4.5 Conclusion | Reflections & Takeaways

### 5 Understanding User Engagement

5.1 Literature Review Methodology

5.2 What Is Engagement?
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2.1 Definitions of Engagement</td>
<td>59</td>
</tr>
<tr>
<td>5.3 Engagement Theory</td>
<td>61</td>
</tr>
<tr>
<td>5.3.1 Levels of Theory</td>
<td>61</td>
</tr>
<tr>
<td>5.3.2 Cognitive, Emotional and Behavioural Engagement</td>
<td>63</td>
</tr>
<tr>
<td>5.3.3 Trait, State or Process Characterisations</td>
<td>64</td>
</tr>
<tr>
<td>5.3.4 Prefixation</td>
<td>65</td>
</tr>
<tr>
<td>5.4 The Ecology of Engagement</td>
<td>65</td>
</tr>
<tr>
<td>5.4.1 The Engaged State</td>
<td>66</td>
</tr>
<tr>
<td>5.4.2 The Engaged Agent</td>
<td>66</td>
</tr>
<tr>
<td>5.4.3 Engaged Interaction</td>
<td>67</td>
</tr>
<tr>
<td>5.5 Measuring Engagement</td>
<td>68</td>
</tr>
<tr>
<td>5.5.1 The Practice of Measurement</td>
<td>69</td>
</tr>
<tr>
<td>5.5.2 Subjectivity-Oriented Approaches</td>
<td>70</td>
</tr>
<tr>
<td>5.5.3 Objectivity-Oriented Approaches</td>
<td>72</td>
</tr>
<tr>
<td>5.6 Design Strategies for Engagement</td>
<td>75</td>
</tr>
<tr>
<td>5.6.1 Ensuring Usability</td>
<td>75</td>
</tr>
<tr>
<td>5.6.2 Managing User Resources</td>
<td>77</td>
</tr>
<tr>
<td>5.6.3 Immersing Users</td>
<td>79</td>
</tr>
<tr>
<td>5.6.4 Engaging Users Emotionally</td>
<td>81</td>
</tr>
<tr>
<td>5.6.5 Inspiring Fun</td>
<td>82</td>
</tr>
<tr>
<td>5.6.6 Enabling Exploration</td>
<td>83</td>
</tr>
<tr>
<td>5.6.7 Supporting Social Connectedness</td>
<td>84</td>
</tr>
<tr>
<td>5.6.8 Enabling Optimal Experience</td>
<td>86</td>
</tr>
<tr>
<td>5.6.9 Promoting Repeated Use</td>
<td>86</td>
</tr>
<tr>
<td>5.7 Discussion</td>
<td>89</td>
</tr>
<tr>
<td>5.7.1 Domains of Knowledge</td>
<td>89</td>
</tr>
<tr>
<td>5.7.2 What's Missing?</td>
<td>91</td>
</tr>
<tr>
<td>5.7.3 Understanding Engagement</td>
<td>93</td>
</tr>
<tr>
<td>5.7.4 The Value of Engagement</td>
<td>96</td>
</tr>
<tr>
<td>5.8 Conclusion</td>
<td>Reflections &amp; Takeaways</td>
</tr>
</tbody>
</table>
8 Results | The Self-Report of Wellbeing in Pregnancy

8.1 The Study Population ................................................. 137
  8.1.1 App Installation by Population ............................... 138
  8.1.2 App Installation by Wellbeing ............................... 141
  8.1.3 Device Type by Population .................................. 142
8.2 Evidence for Feasibility | Women's Engagement ......................... 142
  8.2.1 Evidence for Engagement | Self-Report ......................... 142
  8.2.2 Evidence for Engagement | Protocol Adherence .......... 144
  8.2.3 Evidence for Engagement | Time Spent ....................... 146
  8.2.4 Evidence for Engagement | Population & Wellbeing .... 147
  8.2.5 Evidence for Engagement | Users in Distress ............. 149
  8.2.6 Subjective Engagement | Women's Reflections on the Use of BrightSelf ... 150
8.3 Conclusion | Reflections & Takeaways ............................. 155

9 Discussion & Conclusion .............................................. 157
  9.1 In Summary ......................................................... 157
  9.2 Future Work ....................................................... 160
    9.2.1 Qualitative Insight | Midwives' Reflections ............ 161
    9.2.2 Future Clinical Studies | Midwife Feedback ............. 161
    9.2.3 Analysis of the BrightSelf Corpus ....................... 162
  9.3 Speaking to Complexity ......................................... 164
  9.4 Labelling (Ill and Well) Being .................................. 164
  9.5 Remaking the Self ............................................... 165
  9.6 Our Orientation in Time ........................................ 166
  9.7 Negotiating Care in Extreme Uncertainty .................... 167
  9.8 The Pursuit of Wellbeing ....................................... 168
  9.9 Conclusion ....................................................... 170

Notes ............................................................. 173

Bibliography ..................................................... 189
<table>
<thead>
<tr>
<th>Appendix A</th>
<th>Design Session Structure</th>
<th>249</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix B</td>
<td>Design Work</td>
<td>251</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Wellbeing Data Examples</td>
<td>273</td>
</tr>
</tbody>
</table>
List of Figures

1.1 The Thesis Timeline .............................................. 5
2.1 The Mental Health Spectrum. Adapted From Huppert, 2014 ................. 17
2.2 The 'Heckman Curve.' Adapted From Doyle et al., 2009 ..................... 19
2.3 Reported Symptoms of Perinatal Illbeing. Adapted From Boots Family Trust, 2013 . 20
3.1 The Consolidation of Self ........................................ 26
3.2 The Construal of Experience .................................... 30
3.3 Inter-Self Dissonance ........................................... 34
5.1 The Review Process ............................................. 57
5.2 Engagement by Year ............................................ 58
5.3 An HCI Ecology .................................................. 65
6.1 Recruitment Materials ............................................. 100
6.2 Notes Made by Design Session Participants ............................. 102
6.3 Times of Pronounced Concern .................................. 106
6.4 The Antenatal Care Pathway .................................... 107
6.5 The Antenatal Sampling Protocol ................................ 113
7.1 The BrightSelf System Architecture .................................. 119
7.2 BrightSelf | Design Sketching Examples ............................... 120
7.3 BrightSelf | Animating the Logo at Launch ............................. 121
7.4 BrightSelf | The Recruitment Booklet ................................ 123
7.5 BrightSelf | Momentary & Retrospective Reporting ....................... 125
7.6 BrightSelf | Checking In & Checking Back ............................ 126
7.7 BrightSelf | Feedback Following Self-Report ............................ 127
7.8 BrightSelf | Additional Features .............................................. 130
7.9 BrightSelf | Wellbeing Visualised .............................................. 131
7.10 BrightSelf | iOS Notifications ...................................................... 133
7.11 BrightSelf | The Ideas Machine ...................................................... 134
8.1 Participants by Site .......................................................... 138
8.2 Participants by Age .......................................................... 138
8.3 Participants by Ethnicity ..................................................... 139
8.4 Participants by Marital Status .............................................. 139
8.5 Participants by Employment Status ...................................... 139
8.6 Participants by Education ................................................... 140
8.7 Participants by Number of Children ..................................... 140
8.8 Participants by Baseline EPDS Scores ................................... 141
8.9 Reporting by Week | All Users ............................................... 142
8.10 The Total Number of Reports by User | Arm 1 ......................... 143
8.11 The Total Number of Reports by User | Arm 2 ......................... 143
8.12 Protocol Adherence | Arm 1 .................................................. 144
8.13 Protocol Adherence | Arm 2 .................................................. 144
8.14 Adherence Trends | Arms Compared (LOESS Fit Shown) ............. 145
8.15 Time Spent by Week | All Users ............................................... 146
8.16 Average Session Duration by Week | All Users ......................... 146
8.17 Ideas Machine Usage by Week | All Users .............................. 147
8.18 Reports by User & Age | All Users ........................................ 148
8.19 Baseline EPDS Scores by Ideas Machine Usage | All Users ............. 148
8.20 EPDS Scores by Week | Users Who Provided Multiple Alerts ......................... 149
8.21 Post-Study Survey Response | Ease of Use ............................. 150
8.22 Post-Study Survey Response | Learnability ............................. 151
8.23 Post-Study Survey Response | Women's Needs ......................... 151
8.24 Post-Study Survey Response | The Value of Assessments ............... 152
8.25 Post-Study Survey Response | Engagement .............................. 152
8.26 Post-Study Survey Response | Would Recommend to a Friend ............ 153
8.27 Post-Study Survey Response | Would Repeat the Experience ............ 153
8.28  Post-Study Survey Response | Reporting Frequency . . . . . . . . . . . . . . . 154
8.29  Post-Study Survey Response | Comparing Momentary & Retrospective Reports . . 154
8.30  Perceptions of Reporting Frequency v. Total Number of Reports | All Users . . . . . 155

9.1  The BrightSelf Online Interface | The Midwife's Perspective . . . . . . . . . . . . . . 160
9.2  Women's Risk of Depression by Week | All Users . . . . . . . . . . . . . . . . . 161
9.3  The Wellbeing of Participant No. 212 by Week | Arm 2 . . . . . . . . . . . . . . 162
9.4  The Wellbeing of Participant No. 47 by Week | Arm 2 . . . . . . . . . . . . . . 164
9.5  The BrightSelf Online Interface | The Home Screen . . . . . . . . . . . . . . 165
9.6  Women's Enjoyment by Week | All Users . . . . . . . . . . . . . . . . . . . . . 169
List of Tables

1.1 Critical Realism. Adapted from Grix, 2010 ............................................. 9

4.1 EMA Technologies (Pre-2005) ................................................................. 43

4.2 EMA Technologies (Post-2005) ................................................................. 44

5.1 Engagement & Technology ................................................................. 61

5.2 Engagement & Theory ................................................................. 62

5.3 Engagement & Measurement ................................................................. 69

5.4 Engagement & Application Context ................................................................. 76

6.1 Maternal Participants ................................................................. 101

6.2 Participant Inclusion & Exclusion Criteria ................................................................. 114

8.1 Reporting by Arm ................................................................. 145
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM</td>
<td>Association for Computing Machinery.</td>
</tr>
<tr>
<td>ADHD</td>
<td>Attention Deficit Hyperactivity Disorder.</td>
</tr>
<tr>
<td>CBT</td>
<td>Cognitive Behavioural Therapy.</td>
</tr>
<tr>
<td>CMD</td>
<td>Common Mental Disorder.</td>
</tr>
<tr>
<td>CSO</td>
<td>Clinical Studies Officer.</td>
</tr>
<tr>
<td>DSM</td>
<td>Diagnostic and Statistical Manual of Mental Disorders.</td>
</tr>
<tr>
<td>ECG</td>
<td>Electrocardiogram.</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography.</td>
</tr>
<tr>
<td>ELM</td>
<td>Elaboration-Likelihood Model of Persuasion.</td>
</tr>
<tr>
<td>EMA</td>
<td>Ecological Momentary Assessment.</td>
</tr>
<tr>
<td>EMI</td>
<td>Ecological Momentary Intervention.</td>
</tr>
<tr>
<td>EPDS</td>
<td>Edinburgh Postnatal Depression Scale.</td>
</tr>
<tr>
<td>ESM</td>
<td>Experience Sampling Method.</td>
</tr>
<tr>
<td>ESQ</td>
<td>Engagement Sampling Questionnaire.</td>
</tr>
<tr>
<td>GAD</td>
<td>Generalised Anxiety Disorder.</td>
</tr>
<tr>
<td>GP</td>
<td>General Practitioner.</td>
</tr>
<tr>
<td>GSR</td>
<td>Galvanic Skin Response.</td>
</tr>
<tr>
<td>HCI</td>
<td>Human-Computer Interaction.</td>
</tr>
<tr>
<td>HF</td>
<td>Heat Flux.</td>
</tr>
<tr>
<td>HSM</td>
<td>Heuristic-Systemic Model of Social Information Processing.</td>
</tr>
<tr>
<td>ICL</td>
<td>Imperial College London.</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronics Engineers.</td>
</tr>
<tr>
<td>MBCT</td>
<td>Mindfulness-Based Cognitive Therapy.</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>MBSR</td>
<td>Mindfulness-Based Stress Reduction.</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service.</td>
</tr>
<tr>
<td>OCD</td>
<td>Obsessive Compulsive Disorder.</td>
</tr>
<tr>
<td>OS</td>
<td>Operating System.</td>
</tr>
<tr>
<td>PND</td>
<td>Perinatal Depression.</td>
</tr>
<tr>
<td>PTSD</td>
<td>Post-Traumatic Stress Disorder.</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomised Controlled Trial.</td>
</tr>
<tr>
<td>SBP</td>
<td>Systolic Blood Pressure.</td>
</tr>
<tr>
<td>SDT</td>
<td>Self Determination Theory.</td>
</tr>
<tr>
<td>SMS</td>
<td>Conway &amp; Pleydell-Pearce's Self Memory System.</td>
</tr>
<tr>
<td>TCD</td>
<td>Trinity College Dublin.</td>
</tr>
<tr>
<td>UES</td>
<td>User Engagement Scale.</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom.</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization.</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

This thesis examines the design and feasibility of mobile applications for the self-report of psychological wellbeing in pregnancy. It asks how technology shapes the self-report of wellbeing, how users engage and are engaged in the honest disclosure of mental health concerns, how health professionals might act upon reports of psychological wellbeing, and how technology might contribute to our evolving conception of wellbeing and its pursuit.

1.1 Mental Health & Wellbeing

1.1.1 Wellbeing

At the present moment, 1 in 6 adults living in England meets the diagnostic criteria for a common mental disorder (CMD) (McManus, Bebbington, Jenkins, & Brugha, 2016). Across the United Kingdom (UK), mental illness accounts for nearly half of all illness among those under the age of 65 (Layard et al., 2012), and has a more significant impact on life expectancy than obesity; reducing life expectancy by 7 to 10 years among those with depression, and by 10 to 15 years in the case of schizophrenia (Chang et al., 2011; Layard et al., 2012).

Wellbeing is no longer considered a purely philosophical concern but a valuable resource which allows us to cope with our environment, satisfy our needs, realise our aspirations and live “long, productive, and fruitful” lives (Centers for Disease Control and Prevention, 2017). Wellbeing entails both objective and subjective components, temporal and experiential threads whose patterns are individual and context-specific. Measures of wellbeing are expected not only to articulate the meaning of the concept but to say something of how we should live our lives, and thereby shape our societies, services and technologies.
1.1.2 The Public Health Approach to Mental Healthcare

A public health approach is widely considered the most effective means of improving health and well-being on a population scale (WHO, 2005). Effective treatments, both psychotherapeutic and pharmaceutical, exist for many common mental illnesses (F. Huppert, 2014). Appropriate programs of assessment and avenues of disclosure are required however, to ensure that those in need can access these services.

In 2014, less than 40% of adults living in England with common mental health disorders such as anxiety or depression were in receipt of treatment (Foley, 2013; McManus et al., 2016). Epidemiological surveys of comparable Western countries reveal median treatment rates of 29.3% for depression and 31.6% for Generalised Anxiety Disorder (GAD), in sharp contrast to the treatment of long term physical health problems such as heart disease (78%), hypertension (91%) and diabetes (94%) (Ormel et al., 2008). The World Health Organization's (WHO) most recent estimates reveal similar treatment gaps for schizophrenia (17.8%), Obsessive Compulsive Disorder (OCD) (24.6%), bipolar disorder (39.9%), dysthymia (43.9%), depression (45.4%), panic disorder (47.2%), GAD (62.3%), and alcohol abuse and dependence (92.4%) across Europe (WHO, 2004). Bridging such treatment gaps is one of the primary aims of a public health approach to mental healthcare.

1.1.3 Mental Health in the Perinatal Context

Pregnancy is a pivotal time in many people's lives, and a public health priority given the relationship between perinatal mental illness and long-term adverse health outcomes in women and children. Pregnant women suffering from depression are more likely to engage in unhealthy practices including poor diet, substance abuse and failure to enrol in prenatal care, and are at increased risk of self-harm, suicide and postnatal depression (J. Heron, O'Connor, Evans, Golding, & Glover, 2004; Stewart, 2011). Antenatal depression can also affect fetal development and has been identified as an independent risk factor for children's emotional, cognitive, behavioural and social development through adolescence (Kinsella & Monk, 2009; Stewart, 2011; Pearson et al., 2013; Braithwaite, Murphy, & Ramchandani, 2016). Each case of perinatal depression (PND) results in an estimated cost of £74,000 to the UK's public health system — a total of £8.1 billion for each one-year cohort of birth (A. Bauer, Parsonage, Knapp, Iemmi, & Adelaja, 2014).

Every year, up to 15% of women are diagnosed with depression during pregnancy or within a year of giving birth (Patients Association, 2011). Rates of perinatal depression may be as high as 35% among certain demographic groups (Moses-Kolko & Roth, 2003), and PND may also affect up to 10% of men
Suicide is the leading cause of maternal mortality within a year of birth across the United Kingdom (UK) (Knight et al., 2016). Significantly however, it is estimated that at least 50% of PND cases go undiagnosed (Thurgood, Avery, & Williamson, 2009; NHS Improving Quality, 2015). In the absence of routine screening, up to 75% of women meeting the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for anxiety and depression are not identified (A. O. Coates, Schaefer, & Alexander, 2004; Kingston et al., 2015) and only 1 in 10 women requiring mental healthcare receives it (Bowen, Bowen, Butt, Rahman, & Muhajarine, 2012). Barriers to identifying those in distress and supporting help-seeking behaviours include difficulties differentiating depressive symptoms from the expected mood and somatic changes of pregnancy, stigma, a lack of reassurance that mental healthcare is a normal part of antenatal care, the configuration of health services, characteristics of healthcare providers, insufficient consultation time and the costs of screening practices (Goodman, 2009; Howard et al., 2014; Kingston et al., 2015).

1.1.4 The Self-Report of Wellbeing

The practice of psychiatry, clinical psychology and healthcare more generally is often based in reported experience. Mental health professionals’ knowledge of patients’ wellbeing is typically global and retrospective in nature, informed by the clinical interview and validated screening questionnaires. However, there is growing concern that these methods may neglect the lived experience of health and wellbeing. Several decades of cognitive psychology and behavioural economics research have revealed striking differences between what patients recount of their wellbeing in the moment and looking back over time (Shiffman, Stone, & Hufford, 2008; Ebner-Priemer & Trull, 2009b). If what we express of our wellbeing is coloured by the experience of self-report itself, then the now near-ubiquitous presence of mobile devices in our daily lives may facilitate a fresh perspective.

1.2 Technology for the Self-Report of Wellbeing

In the context of the UK’s National Health Service (NHS), mental health assessments in pregnancy are currently carried out verbally and using paper-based questionnaires completed in the waiting rooms of antenatal clinics. Mobile devices have the potential to facilitate the remote screening and monitoring of wellbeing and depression throughout the antenatal period, extending care to under-served and at-risk populations, enabling timely assessment and intervention, gathering ecologically valid and longitudinal data, overcoming stigma, supporting honest disclosure, and fostering trust between women and midwives.
However, we know little about the feasibility of deploying these systems in practice, including whether these technologies can be designed to engage women and midwives, gather meaningful data, manage high-risk cases, interface with existing clinical practices, generate manageable workloads, meet the needs of patients and healthcare providers or overcome ethical and privacy concerns. To date, millions of women have installed thousands of mobile applications in the hope of supporting a healthy start to life.\textsuperscript{4} However, to the best of the author’s knowledge, none have yet been designed to support the self-report of mental health and wellbeing in combination with antenatal care nor have aimed to advance our collective understanding of perinatal wellbeing. Few studies have examined users’ longitudinal engagement with mobile technologies for the self-report of mental health and wellbeing in general (Shiffman et al., 2008).

1.2.1 Engaging Women & Midwives

In 2017, 94% of UK adults report owning a mobile phone, and 76% a smartphone (Ofcom, 2017). On average, we tap, swipe and click our phones 2,617 times a day (dscout, 2016), 1 million interactions a year, totalling more than 3 hours a day spent with a mobile device (Meeker, 2017). Mobile applications are in constant competition for our time and attention.

The engagement of patients and their willingness to run screening or clinical monitoring applications on their personal devices is a key factor in the success of mobile health technologies (Torous et al., 2014). In qualitative studies, women report that they refrain from initiating discussions with professionals about their mental health, or provide inaccurate responses to screening scales, due to discomfort, stigma and uncertainty with respect to the ‘normal’ emotional experience of pregnancy (Byatt et al., 2013; Kingston et al., 2015). Mobile technologies may have the capacity to overcome these barriers to care — if designed to support the needs, values and motivations of women and midwives. Little design research has been conducted in the perinatal context to date, however.

1.3 The Research Question

Through collaboration with public health researchers, pregnant women and a variety of health professionals including midwives, this thesis seeks to establish a case for the feasibility of deploying mobile applications for the self-report of psychological wellbeing and depression in antenatal care. This is a highly complex context for research and design, in light of which, this thesis aspires to address the question;
Can an appropriately designed mobile application engage women in the self-report of wellbeing during pregnancy?

1.4 Thesis Structure

This thesis traces an idea from concept to reality, by means of literature review (Chapters 1-5), theory development (Chapters 3, 5), design research (Chapters 6, 7), technology and protocol design (Chapters 6, 7), clinical deployment, and data analysis (Chapters 8, 9) (See Fig 1.1).

Chapter 2 examines the current state of wellbeing science, including the conception, theory and measurement of wellbeing and the role of technology in its pursuit. This chapter also describes and motivates the public health approach to perinatal mental healthcare. Chapter 3 turns to the topic of self-report; whether we might expect a richer understanding of wellbeing to arise from data, the nature of experience and reflection with respect to multiple time-frames, and motivations for the adoption of Ecological Momentary Assessment (EMA) methodologies. A framework of multiple self-concepts is presented as a means to support the design of self-report technologies and analysis of the data they produce. Chapter 4 charts the design space for self-report technologies, including the methodological, technological and human factors which shape their development and use. Chapter 5 details the findings of a systematic literature review of the conception, theory, measurement and design of user engagement across the Computer Science and Human-Computer Interaction (HCI) literatures, providing the epis-
temological framing required to support a coherent understanding of feasibility and design in terms of engagement.

Chapter 6 describes the experience of pregnancy and perinatal care as recounted by 38 women and health professionals during design research. These participants’ insights informed the development of a randomised controlled trial protocol for the longitudinal clinical deployment of a mobile application during pregnancy, and the chapter concludes with description of this methodology for an antenatal feasibility study. Chapter 7 continues analysis of these sessions, detailing the iterative design and evaluation of a mobile application and online platform for the self-report of psychological wellbeing during pregnancy; BrightSelf. Chapter 8 presents detailed mixed-methods analysis of 355 women's engagement with BrightSelf over the course of 9 months following the deployment of this system in antenatal care, and discusses the implications of these findings for the design and feasibility of mental health screening technologies. Chapter 9 concludes with the author's reflections on the role of mobile devices for the self-report, and pursuit, of wellbeing.

Chapters 6, 7 and 8 describe the results of an interdisciplinary and international collaboration between HCI and public health researchers. José Marcano-Belisario, PhD student at the School of Public Health at Imperial College London (ICL), participated in the conception of the BrightSelf project, helped to organise and attended multiple design sessions, established contact with NHS antenatal clinics across the UK and led the preparation of the ethics submission for the clinical study. Martin Cohn and Nikolas Mastellos, post-doctoral researchers at ICL, fulfilled the team's medical and ethical responsibilities throughout the clinical study, contacting the appropriate clinical teams following alerts for depression, self-harm and suicide risks, responding to midwives’ queries, and managing the collection and secure offline storage of baseline data. Cecily Morrison, of Microsoft Research Cambridge, took part in the BrightSelf project's conception while also helping to organise and conduct design sessions in the Cambridge area. Marguerite Barry, post-doctoral researcher at the School of Computer Science and Statistics, Trinity College Dublin (TCD), travelled to the UK on several occasions to support the conduct of design sessions, assisted in the transcription of health professionals’ reflections, and helped shape the design of BrightSelf as well as our ethical reflections on the project. Bérenger Arnaud, also a post-doctoral researcher at TCD, provided invaluable advice with respect to the development of the BrightSelf online platform. Finally, Gavin Doherty and Josip Car supervised each stage of the BrightSelf project, on behalf of Trinity College Dublin and Imperial College London respectively.5
1.5 An Overarching Methodology | The Epistemology of Design

“Because we are in the world, we are condemned to meaning.” Merleau-Ponty, 1945

Conceptually, the work of this thesis is positioned within HCI and design thinking traditions. It embodies an epistemology of critical realism. This is a highly complex context for research and design, combining the individual significance of wellbeing during pregnancy, the real-world and long-term use of technology by an at-risk user group, a public health system’s need to efficiently distribute resources, midwives’ diverse work practices, social expectations, societal stigma, and researchers’ motivations for data collection. Designing any technology capable of respecting these constraints is a ‘wicked problem’ as first defined by Rittel & Webber in 1973 (Rittel & Webber, 1973).

As the work of this thesis spans several disciplines and employs multiple methodologies, each chapter provides its own methodological outline where appropriate.

1.5.1 Designing for Complexity | A Wicked Problem

Microsoft released the first version of Word for Windows in 1989. Three decades later, computing technologies have emerged from the workplace to inhabit almost every corner of our lives. This is a change to the adjacent possible. Technology is no longer seen as a lever for efficiency but for experience, leisure, social connectedness, health, and wellbeing (Bardzell, Bardzell, Pace, & Karnell, 2008). HCI research embodies this trend. When we speak of technology we speak not of reified objects but of an interplay between behaviour, cognition and emotion, between choices made by designers and in turn by users, of meaning and action, of the bounds of context, and of experience brought to interaction in the form of unique selves. Designers are tasked with untangling complexity “on multiple levels simultaneously - as technological artefacts, social facts, and cultural narratives” (Boehner, DePaula, Dourish, & Sengers, 2005). This is a forward looking activity, which points “to the openness and un-finalizability of a world that, though already half-designed, is always becoming” (McCarthy & Wright, 2004). HCI research often embodies pragmatist and phenomenological perspectives which reflect felt experience and meaning.

1.5.2 Design as Research | Towards What Works

Bruce Archer defined design as the “area of human experience, skill and understanding that reflects man’s concern with the appreciation and adaptation of his surroundings in the light of his material and spiritual needs” (Archer, 1979), Victor Papanek as “the conscious and intuitive effort to impose mean-
ingful order” (Papanek, 1972), and Herbert Simon as an approach to problem solving which strives to ‘satisfice’ rather than ‘optimise’, “shaping and squeezing” real-world problems to requirements (Simon, 1969). Common to these perspectives is a recognition of the complexity of ‘wicked problems’ and the need to realise ‘best possible’ solutions. Proponents of design thinking\(^\text{11}\) construe design as a process for the resolution of ambiguous and large-scale problems. HCI emphasises the non-determinacy of design and the importance of process in building upon the experiences of users, through participatory, user-centred, service and human-centred approaches (Friess, 2010).

1.5.3 Critical Realism | An HCI Epistemology

This thesis approaches the construction of knowledge through an ontological and epistemological perspective of critical realism. Design researchers are essentially interested in establishing what works and why. HCI research challenges the illusion of closed systems implied by the traditional dichotomy between subjective (interpretivist) and objective (positivist) research methods (Dickens, 2004). Quantitative methods can provide evidence for the feasibility and efficacy of mental health technologies by distinguishing the merits of specific solutions or design choices. This is “knowledge that exists externally to the researcher’s interpretive position” (Easton, 2010). Qualitative methods, on the other hand, recognise the complex, interconnected and changing network of personal, cultural, social and systemic factors in which design research is situated (Sayer, 2010) and can provide the knowledge to understand why one system works rather than another, as well as the insight required to support the action of design. This socially constructed knowledge embodies an interpretivist perspective. Designers, like the social workers of whom Carolyn Oliver writes, are required to “balance respect for individual meaning-making with evidence to test that meaning-making for its correspondence to an external reality” (Oliver, 2011). This requires a coherent epistemological account for methods derived from multiple disciplinary origins.

Critical realism (also known as post-positivism) is an epistemological stance which “marries the positivist’s search for evidence of a reality external to human consciousness with the insistence that all meaning to be made of that reality is socially constructed,” accepting that social constructions themselves “constitute what we know as the reality of our social worlds” (Oliver, 2011) by employing pragmatic, constructivist and positivist epistemologies to generate fundamentally constructivist explanations (Sayer, 2010).\(^\text{12}\)

Critical realism posits a hierarchical stratification of reality according to which the physical, biological and social combine to create the world we experience (Dickens, 2004). This ontology permits
Critical Realism supports sense-making and the identification of causality, activities central to design, by seeking to explain relationships “between experiences, events and mechanisms …questions of ‘how and why’” (Jeppesen, 2005). The tenets of critical realism inscribe a middle-ground between paradigms of positivism and interpretivism (See Table 1.1).

### 1.6 Contributions

The primary contributions of this thesis include:

- A contribution to knowledge in the form of an enhanced understanding of design to support wellbeing in the context of antenatal mental healthcare; informed by qualitative analyses of the needs, values and motivations of women and health professionals, grounded in knowledge of the conception, theory and measurement of wellbeing, and realised in the design of BrightSelf.

- An understanding of the design of self-report technologies sensitive to the importance of how we both live and think about experience with respect to multiple time-frames; articulated in the form of a framework of multiple self-concepts of a sufficiently simple character to function as a
practical heuristic and sense-making tool, linked to the features of the design space for self-report technologies, and expressed in terms of implications for design in the context of pregnancy.

- An improved understanding of the conception, theory, measurement and design of user engagement; informed by systematic literature review, augmented by women’s and health professionals’ reflections on the use of technology during pregnancy, actualised in the design of a protocol for the self-report of antenatal wellbeing and depression, and employed to support analysis of the results of a feasibility study with a randomised controlled trial design — ultimately contributing initial evidence for the design and feasibility of mobile applications to engage women in the self-report of wellbeing and depression, extend care to under-served and at-risk groups, enable longitudinal, momentary and retrospective data collection, overcome stigma, support disclosure, and foster trust between patients and health professionals.

This thesis resulted in a new understanding of the feasibility and design of mobile technologies for public health screening and research.
Chapter 2

Wellbeing & Mental Health

For centuries, the pursuit of the life well-lived has been a defining feature of the human experience. Today, research shows that people with high levels of wellbeing are more sociable, generous, creative, active, tolerant, healthy, altruistic, productive, learn more effectively and live longer (Shah & Marks, 2004; F. A. Huppert & So, 2013). This chapter provides an overview of the current state of wellbeing science, including the search for a definition, what positive psychology and public health research tells us about wellbeing, and the emerging role of technology in its pursuit.

2.1 Defining The Life Well-Lived

Wellbeing has been described as “intangible, difficult to define and even harder to measure” (Thomas, 2009). Yet, how we define wellbeing inevitably shapes the policies, services and technologies implicated in its pursuit. The following section traces the conceptual history of wellbeing including descriptions of health and quality of life in medicine, wellness in counselling, and functioning, emotion, happiness and flourishing in psychology.

2.2 Wellbeing as The Absence of Suffering

Prior to World War II, health was viewed primarily in terms of the absence of disease and disability (Cooke, Melchert, & Connor, 2016). In 1937, a working group planning the UK’s national health service wrote:

“Health must come first: the mere state of not being ill must be recognised as an unacceptable substitute, too often tolerated or even regarded as normal. We must, moreover, face the fact that while immense study has been lavished on disease, no-one has intensively
studied and analysed health, and our ignorance of the subject is now so deep that we can hardly claim scientifically to know what health is.” Political and Economic Planning, 1937

The constitution of the World Health Organization (WHO) on its founding in 1946 put forward a new definition of health as “a state of complete physical, mental, and social well-being and not merely the absence of disease and infirmity” (WHO, 1946). The proceeding half-century saw many attempts to study and define the component characteristics of positive health and wellbeing (Gallagher, Lopez, & Preacher, 2009; Dodge, Daly, Huyton, & Sanders, 2012; Cooke et al., 2016).

2.3 Wellbeing as Functioning

The first definition of mental health is typically attributed to Jahoda, who articulated six features of positive functioning; an individual’s attitudes towards his own self, self-actualization, integration, autonomy, perception of reality, and environmental mastery (Jahoda, 1958; F. A. Huppert & So, 2013). This framing evokes the eudaimonic component of Aristotle’s description of the ‘good life,’ pertaining to the fulfilment of potential, optimal functioning and the realisation of one’s ‘true nature’ (Lent, 2004; F. Huppert, 2014).

Subsequent definitions of wellbeing in this vein included Ryff’s model of psychological wellbeing in terms of self-acceptance, positive relations with others, autonomy, environmental mastery, purpose in life and personal growth (Ryff, 1989), Keye’s model of social wellbeing as social integration, contribution, coherence, actualization and acceptance, Antonovsky’s association of wellbeing with a sense of coherence, an enduring attitude through which life is seen as comprehensible, manageable and meaningful (Antonovsky, 1993), and Gough et al’s description of wellbeing in terms of “what people are notionally able to do and to be, and what they have actually been able to do and to be” (Gough & McGregor, 2007).

The WHO currently defines mental health as “a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community” (WHO, 2005).

2.3.1 Wellbeing as Wellness

Descriptions of ‘wellness’ often bear resemblance to definitions of optimal functioning. This term is more often employed within the counselling literature to describe a lifestyle of personal improvement, a way of life oriented towards optimal wellbeing or an optimal state of wellbeing (Cooke et al., 2016).
Dunn first defined wellness as “an integrated method of functioning” oriented towards maximising an individual’s potential according to a “continuum of balance and purposeful direction” (Dunn, 1961 as cited in Cooke et al., 2016). The WHO defines wellness as “the optimal state of health of individuals and groups” related to “the fullest potential of an individual physically, psychologically, socially, spiritually and economically, and the fulfilment of one's role expectations in the family, community, place of worship, workplace and other settings” (B. Smith, Kwok, & Nutbeam, 2006).

2.4 Wellbeing as Subjective Experience

The hedonic\(^8\) tradition emphasises the subjective experience of wellbeing (Gallagher et al., 2009).

2.4.1 Wellbeing as Emotion

Many consider wellbeing to entail an emotional component. Bradburn’s description of psychological wellbeing in 1969, as the degree to which an individual “has an excess of positive over negative affect,” is perhaps the most candid articulation of this perspective (Bradburn, 1969). While reductive, this is not an uncommon interpretation of wellbeing. The Positive and Negative Affect Schedule (PANAS), which consists of two 10-item scales for positive and negative affect, is one of the most frequently employed measures in wellbeing research (Watson, Clark, & Tellegen, 1988; Gallagher et al., 2009).

The Subjective Happiness Scale (SHS) asks respondents to characterise themselves in terms of global subjective happiness (Lyubomirsky & Lepper, 1999). However, the relationship between happiness and wellbeing lacks consensus. Although there have been attempts to align the concept with more complex interpretations of wellbeing, happiness has also been described as “an awkward construct that hides the true, complex, nature of human flourishing” (Dodge et al., 2012).

2.4.2 Wellbeing as Appraisal

Self-report is the methodology of subjective wellbeing, and many measures encourage broad appraisal. Life Satisfaction is “by far the most widely used” measure of wellbeing, often assessed through a single question such as “All things considered, how satisfied are you with your life as a whole these days?” (F. Huppert, 2014).\(^10\) Cantril’s 1965 ‘Ladder of Life’ is one of the earliest and most influential of these scales. It asks “Please imagine a ladder with steps numbered from zero at the bottom to ten at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you personally say you stand at this time?” (F. Huppert, 2014). More recently, Diener defined Subjective Wellbeing (SWB) as the general
evaluation of one’s quality of life, comprising high life satisfaction, positive levels of pleasant emotion, and low levels of negative mood (Diener, 2009).

The term Quality of Life (QoL) is often used interchangeably with life satisfaction and wellbeing — a concept defined by the WHO as “an individual’s perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns” (The WHOQOL Group, 1998). Shin & Johnson’s description of happiness in 1978 as “a global assessment of a person’s quality of life according to his own chosen criteria” underscores the subjective quality of assessment as well as the intersectional nature of these various constructs (Shin & Johnson, 1978).

2.5 Wellbeing as Flourishing

The most recent developments in the description of wellbeing have been the result of positive psychology research, a discipline which has become synonymous with the term ‘flourishing’ (Dodge et al., 2012). Flourishing has been defined as “the experience of life going well,” “a high level of mental wellbeing” and “a combination of feeling good and functioning effectively” (F. A. Huppert & So, 2013).

The PERMA model, an extension of Seligman’s original model of ‘authentic happiness,’ posits positive emotion, engagement, relationships, meaning and accomplishment as “the best approximation of what humans pursue for their own sake” (Seligman, 2012). Diener, a proponent of life satisfaction measures, also recently supplemented this approach to the evaluation of wellbeing with scales for positive and negative experience as well as flourishing. Diener et al’s measure of flourishing encompasses purpose in life, positive relationships, engagement, competence, self-esteem, optimism, and a contribution towards the wellbeing of others (Diener et al., 2010).

2.6 Technologies for Wellbeing

Research suggests that 50% of our wellbeing is linked to our genetic make-up and early environment, 10% to our circumstances, and 40% to our outlook and activities (Lyubomirsky, Sheldon, & Schkade, 2005). These are the domains in which technologies may serve as ‘tools for the good life’ (Schüll, 2016).

2.6.1 Technology, Genes & The Early Environment

It has been claimed that a “disposition for well-being is as heritable as height or intelligence” (Kahneman, 2011). This ‘set-point’ for wellbeing is a result of our parents’ influence; a combination of our
genes and early environment (Lykken, 1999; Shah & Marks, 2004; Lyubomirsky et al., 2005). Genetic expression is subject to environmental factors, meaning that a child’s early environment, including the antenatal period, plays a pivotal role in the expression of positive or negative predispositions (Shah & Marks, 2004). Parental wellbeing has significant and long-lasting intergenerational effects.

2.6.2 Technology & Our Circumstances

Wellbeing is typically sensitive to life’s circumstances and major events, including age, unemployment, education, and marriage (Centers for Disease Control and Prevention, 2017). Material disadvantage, including poor housing, inability to afford a one-week holiday and difficulty making ends meet, is strongly correlated with lower levels of subjective wellbeing (F. Huppert, 2014). Once basic needs for peace, economic security, food, shelter, income, and access to modern conveniences are met however, further improvements to our circumstances tend to have little effect on wellbeing (Shah & Marks, 2004). Although the UK’s economic output has more than doubled over the last 40 years, reported life satisfaction has remained flat (Shah & Marks, 2004). These findings suggest that we quickly adapt to changes in our circumstances.

Gilbert & Wilson coined the term ‘miswanting’ to describe choices made based on false assumptions concerning their impact on happiness and wellbeing (Gilbert & Wilson, 2000). The purchase of a new car is a quintessential example. We may typically spend a significant period of time contemplating this decision and the joy it might bring. In daily life, however, we may rarely pay attention to the pleasure of driving, and therefore overestimate the significance of this choice in comparison to those which better maintain our attention in the long run, such as learning to play the piano or taking up tennis. This is a focusing illusion which underlines the fact that “nothing in life is as important as you think it is when you are thinking about it” (Kahneman, 2011). 

Alternatively, we might think of wellbeing as ‘time well spent.’ How we choose to spend our time does not always reflect that which we value however. Technologies capture much of our time and attention and are potentially well-placed to ‘make the wellbeing choice the easy choice’ (Shah & Marks, 2004). Although technologies are “typically associated with new choices, connections, and forms of self-expression,” they can also “function to narrow choices, disconnect, and gain exit from the self” (Schüll, 2012). The many advertisements we see in a day, for example, provide no shortage of “training in desire and striving,” rather than encouraging us to value spending our time in ways which better support lasting wellbeing (Calvo & Peters, 2014).
Research suggests that we do not adapt to our experiences to the same extent that we do our income and possessions (Shah & Marks, 2004). There is “abundant evidence that the skills for sustainable happiness or wellbeing can be learnt” (F. Huppert, 2014).18

Our wellbeing is linked to our goals and values.19 Goals professed at the age of 18, for example, appear to have a significant effect on life satisfaction 20 years later. In 1976, 12,000 students beginning higher education rated their desire to be ‘very well-off financially’ on a 4-point scale. When the 597 medical professionals within the sample were recontacted in 1995, each additional point correlated with an increment of more than $14,000 in income. Those who rated wealth highly and achieved that goal were significantly more satisfied than average. Those who desired financial success and did not achieve it, however, were significantly more dissatisfied (Kahneman, 2011).20

Wellbeing is also a social phenomenon, “rooted in our broader communities” (Shah & Marks, 2004). Studies have found that wellbeing is most strongly correlated with partner and family relationships, that trust in public institutions is linked to higher life satisfaction, that spending money on others leads to greater happiness, that a sense of belonging to one’s community has a larger effect on life satisfaction than trebling household income, and that membership of a club or society after the age of 50 has a similar effect on the risk of mortality as exercise (Shah & Marks, 2004; F. Huppert, 2014; Steffens, Cruwys, Haslam, Jetten, & Haslam, 2016).

Finally, wellbeing is strongly correlated with our intentional activities, including physical activity,21 sufficient sleep, and the consumption of fresh fruit and vegetables (F. Huppert, 2014). These are physical and cognitive behaviours which technology can support or inhibit. Positive computing researchers recently proposed a framework of nine wellbeing constructs (positive emotions, motivation and engagement, self-awareness, mindfulness, resilience, gratitude, empathy, compassion and altruism) for which there exist evidence-based strategies, including positive rumination, intrinsic and extrinsic rewards, Cognitive Behavioural Therapy (CBT), meditation, Mindfulness-Based Stress Reduction (MBSR) and Cognitive Therapy (MBCT) methods, gratitude journalling, role playing, perspective taking, and emotion-recognition training (Calvo & Peters, 2014).22

Mobile applications have been designed to support health and wellbeing by offering mindfulness techniques, improving sleep and rest skills, facilitating healthier lifestyles, providing health-related information, and serving as tools for the management of anxiety, Post-Traumatic Stress Disorder (PTSD) and depression, for example (Donker et al., 2013; Calvo & Peters, 2014; Rennick-Egglestone et al.,...
Mobile technologies have also allowed users to record and monitor symptoms of mental health and wellbeing (Servia-Rodríguez et al., 2017), Parkinson's disease (Vega et al., 2018), PTSD (Larsen, Christiansen, & Eskelund, 2017), asthma (Buonocore, Rocchio, Roman, King, & Sarrafzadeh, 2017), pain (A. T. Adams et al., 2018), stress and sleepiness (Paruthi et al., 2017), bipolar disorder (Matthews et al., 2015), depression (Matthews & Doherty, 2011), anxiety (Topham, Caleb-Solly, Matthews, Farmer, & Mash, 2015), mood (Van der Velden & Machniak Sommervold, 2016), menstruation (Epstein et al., 2017) and more.

Few technologies, however, have yet been designed to facilitate interaction between patients, healthcare professionals and health services, despite the fact that a public health approach is widely considered the most effective means of “improving mental health,” as well as “lowering the personal and social costs of mental ill-health” (WHO, 2005).

2.7 Fostering Wellbeing | The Public Health Approach

Wellbeing, whether the absence of suffering, feeling good, functioning well or judging life positively, is a spectrum (See Fig 2.1). A public health approach aims to shift an entire population towards positive mental health, compelled by the epidemiological evidence for doing so (F. Huppert, 2014).

This objective hinges upon effective programs of assessment and intervention.

How public health services operationalise wellbeing is reflected in mental health screening programs and clinical care pathways. Effective screening programs enable services to identify those in distress, efficiently distribute resources, and take appropriate action. These are, however, measures of illbeing (distress, depression and anxiety) rather than wellbeing (flourishing). At present, measures
of wellbeing are largely constrained to the domain of research and “resistance to prioritizing positive outcomes remains high in the field of health” (F. Huppert, 2014).

Researchers and clinicians have repeatedly called for the reorganisation of care to support a more proactive approach to mental healthcare, enabling earlier support and intervention (Kazdin & Blase, 2011). Multidimensional measures of wellbeing may shed light on the relationship between flourishing and mental illness, revealing “how the science of well-being can most effectively aid the practice of clinical psychology” (Gallagher et al., 2009). First however, we must “study flourishing in its own right, and not as the mere absence of mental disorder” (F. A. Huppert & So, 2013).

This is a conceptual shift which recognises patients as ‘co-creators' of their health and wellbeing (Shah & Marks, 2004) while respecting the expertise and responsibilities of health professionals. Research suggests that patients show improved maintenance of healthy behaviours, better medication adherence, improved physical and mental health, report greater satisfaction, and make fewer healthcare visits when medical professionals are patient-centred, empathetic and involve patients in decision-making (Williams, Frankel, Campbell, & Deci, 2000; Shah & Marks, 2004).

2.7.1 Pregnancy | An Opportunity to Improve Population Health

Pregnancy is a unique and compelling context for the study of wellbeing, and a critical stage for population health. A healthy start to life has been shown to “greatly enhance …functioning in school, with peers, in intimate relations, and with broader connections with society” (WHO, 2005). Pregnant women suffering from depression are more likely to engage in unhealthy practices including poor diet, substance abuse and failure to enrol in prenatal care, and are at increased risk of self-harm, suicide and postnatal depression (J. Heron et al., 2004; Stewart, 2011). Antenatal depression can also affect fetal development and has been identified as an independent risk factor for a child’s emotional, cognitive, behavioural and social development through adolescence (Kinsella & Monk, 2009; Stewart, 2011; Pearson et al., 2013; Braithwaite et al., 2016). Cost-benefit analysis suggests that investment in the first years of life repays itself ‘many times over’ (See Fig 2.2) (Shah & Marks, 2004; Doyle, Harmon, Heckman, & Tremblay, 2009).

Perinatal depression (PND) affects up to 15% of women during pregnancy or within one year of giving birth in the UK (Patients Association, 2011), and suicide is the leading cause of maternal mortality (Knight et al., 2016). During a recent UK survey, almost 1 in 3 of 1,500 women reported experiencing some level of depression or anxiety during pregnancy (Boots Family Trust Alliance, 2013). More than 1 in 5 (22%) reported suicidal thoughts. A fifth of women described recognising that they were
unwell during pregnancy, and a further 1 in 10 reported doing so in hindsight. Forty percent of those in a relationship reported that their partner had experienced anxiety or depression (Boots Family Trust Alliance, 2013).

Meta-analysis points to antenatal depression as the strongest predictor of postnatal depression alongside a history of depression, antenatal anxiety, stressful life events, negative cognitive attributional style, low self-esteem, low social support and low income (Leigh & Milgrom, 2008). The point prevalence of antenatal depression is significant, at 7.4% in the first trimester, 11.4 to 12.8% in the second, and 13.1 to 14.8% in the third (A. Bauer et al., 2014). Treating depression during these stages can therefore reduce the likelihood of developing postnatal depression, prevent more severe forms of the condition, reduce its intergenerational impact, and improve a woman’s overall health status (Stewart, 2011; Knight et al., 2016).

Midwifery clinics across the UK screen for depression through conversations with women and the completion of questionnaires on paper in waiting rooms. Ninety-six percent of UK midwives report asking women about their psychological wellbeing at their first appointment (Boots Family Trust Alliance, 2013). Two-thirds of professionals report employing the Whooley Questions, two-thirds the Edinburgh Postnatal Depression Scale (EPDS) (including 80% of health visitors), and 12% the Hospital Anxiety and Depression Scale (HADS) (primarily family nurse practitioners).28

However, only 1 in 10 women recall being asked about their mental health, and almost half report that they were never told about the possibility of mental health problems. It is estimated that at least 50% of PND cases go undiagnosed (Thurgood et al., 2009; NHS Improving Quality, 2015). Rates of
PND may be as high as 35% among certain demographic groups (Moses-Kolko & Roth, 2003), and PND may also affect up to 10% of men (NHS Choices, 2016; National Childbirth Trust, 2016).

The emotional experience of pregnancy can make it difficult for women to differentiate depressive symptoms from expected mood and somatic changes. Many of the 18 symptoms of emotional illbeing described by women in a recent survey do not feature in formal diagnostic criteria such as the ICD-10 for example (See Fig 2.3). Depression is not the only form of emotional distress which can affect parents at the time of pregnancy, and recent research has highlighted the need to recognise a greater variety of symptoms and disorders including anxiety and PTSD (R. Coates, de Visser, & Ayers, 2015).

Depression during pregnancy is marked by an unwillingness to seek help at what parents believe should be a happy time (Moore, Ayers, & Drey, 2016). Less than a fifth of women report being completely honest with professionals and a third report never telling a health professional that they had felt unwell (Boots Family Trust Alliance, 2013). Mobile technologies may be able to play a role in overcoming these barriers to care, by increasing women’s insight with respect to their own mental health and wellbeing, facilitating more effective screening programs, and extending avenues for disclosure and support to those in distress.

2.8 Conclusion | Reflections & Takeaways

This overview of wellbeing both motivates and informs the work of this thesis. How we conceive of mental health and wellbeing shapes not only how we measure these concepts but the distribution of care
on national and international scales. This thesis embraces a multi-faceted interpretation of wellbeing, sensitive to the assumptions of multiple models of care and conceptions of mental health and illness with distinct disciplinary origins — overlapping, at times counter-posed, and invariably subject to the double hermeneutic.

As subsequent chapters shall describe in greater detail, this orientation culminated in the adoption of multiple distinct measures of wellbeing; a validated clinical scale for depression, self-harm and suicide, as well as visual analogue scales for mood, rest, worry, enjoyment and energy. This combination ensures a broad conceptual stance spanning a spectrum of ill and wellbeing, providing multiple avenues for disclosure, permitting multidimensional analysis of the correlates of wellbeing, and enabling exploration of women’s and health professionals’ needs with respect to the assessment of wellbeing and depression. This approach is both pragmatic and forward-looking, entailing data capture to support both existing clinical practices contingent upon targeted forms of understanding, and future research and design efforts subject to the shared interpretation of health and wellbeing by women and health professionals. This range of measures further facilitates insight into the experience of wellbeing with respect to multiple time-frames, an important consideration in the design of self-report technologies, to which Chapter 3 is addressed.

In summary, wellbeing science points to the pivotal importance of the antenatal period for population health. Parental wellbeing has significant and long-lasting intergenerational effects. Up to 15% of women are diagnosed with perinatal depression every year, and suicide is the leading cause of maternal mortality within the UK. Effective mental health screening programs enable public health services to extend care and support to those in distress. At present however, it is estimated that at least 50% of PND cases go undiagnosed. Mobile technologies have the potential to facilitate the self-report of mental health and wellbeing throughout the antenatal period, overcoming stigma, supporting disclosure, enabling more effective screening programs, facilitating access to care, improving population health, and contributing to our knowledge of the spectrum of wellbeing during pregnancy.
Chapter 3

Understanding Subjective Experience

To measure wellbeing is to make sense of the contents of consciousness; thoughts, feelings and sensations. Most modern psychotherapies promote awareness of thoughts, emotions and behaviour (Calvo & Peters, 2014). And, much of our knowledge in psychology, HCI and design rests “on what people tell us” (L. F. Barrett & Barrett, 2001).

3.1 The Temporality of Wellbeing

Self-report is in turn a bracketing of subjective experience. This is not a “direct ‘pipeline’ into consciousness” (Christensen, Barrett, Bliss-Moreau, Lebo, & Kaschub, 2003). Wellbeing is more than a positive feeling, the capacity to function effectively, or life satisfaction. Its assessment is subject to “questions with different timeframes, including ongoing experiences, recent experiences, and typical experiences” (Felicia Huppert writing in Calvo & Peters, 2014). We must choose what to assess and how to go about it. The temporality of experience, past, present and future, matters.

3.2 Articulating Experience

“From moment to moment the hedonimeter varies; the delicate index now flickering with the flutter of the passions, now steadied by intellectual activity, now sunk whole hours in the neighbourhood of zero, or momentarily springing up towards infinity.” Edgeworth, 1881

In 1881, Irish economist Francis Edgeworth first wrote of the hypothetical ‘hedonimeter,’ an in-
instrument whose scale would reflect the pleasure or pain of its user’s every moment. While the existence of such a device would undoubtedly evoke no shortage of feeling, many recent technological developments have facilitated the articulation of our inner experience. We are increasingly engaged in practices of self-report and self-expression, by means of social networks, online forums and mobile applications for behaviour change, self-tracking, relationships and more. We are invited to ask whether, and how, it is possible to develop a richer understanding of our subjective experience from self-reported data.

3.3 Ecological Momentary Assessment

The immersion of mobile devices in our stream of consciousness provides unprecedented opportunities for psychology, HCI and wellbeing research, as well as mental health assessment and intervention. Ecological Momentary Assessment (EMA) is the methodology of self-report in the moment of daily life (Shiffman et al., 2008). Although the term was first introduced in 1994 (Stone & Shiffman, 1994), no individual can be credited with the development of a method with antecedents and applications in the fields of psychology, ethology, sociology, anthropology, HCI and mental health care (Stone, Shiffman, & DeVries, 1999). The method’s history is dotted with studies of note (Wilhelm & Perrez, 2013). EMA possesses three key methodological advantages; the circumvention of retrospective bias, temporal and longitudinal validity, and ecological validity (Shiffman et al., 2008; Ebner-Priemer & Trull, 2009b). The Experience Sampling Method (ESM), Ambulatory Assessment (AA), Diary Methods and Mobile Living Labs all refer to this category of method (Broens, ter Hofte, & Ebben, 2009; Scollon, Prieto, & Diener, 2009), and have more recently been employed by HCI researchers to facilitate insight into users’ experiences, and to support the design of health and wellbeing technologies.

3.4 Making Sense of Experience in HCI

“In ontological designing, we are doing more than asking what can be built. We are engaging in a philosophical discourse about the self — about what we can do and what we can be.” Winograd & Flores, 1986

Design is the elicitation, interpretation, and creation of experience. These are processes contingent upon self-report. More often than not, this results in a retrospective framing, rendering design research a process of ‘retrospective reversal’ (Conway, Loveday, & Cole, 2016). The more technologies mediate our experiences, the more difficult it can become to extract meaningful insights with respect to their use.
Several HCI researchers have proposed models of experience designed to support understanding. The early work of Forlizzi & Battarbee featured a distinction between 'experience,' 'an experience,' and 'experience as story' in the context of a transition from 'cognitive' to subconscious awareness (Forlizzi & Ford, 2000; Forlizzi & Battarbee, 2004). The User Experience Lifecycle Model, ContinUE, pursues a similar distinction between phases of the product lifecycle; experience before, during, and after use, and recurring (Pohlmeyer, Hecht, & Blessing, 2009). Karapanos et al. model users’ experience as a form of technological adoption, according to which increased familiarity, functional dependency and emotional attachment shape the experience of use over time (Karapanos, Zimmerman, Forlizzi, & Martens, 2010). They write that “while early experiences seemed to relate mostly to hedonic aspects of product use, prolonged experiences became increasingly … tied to aspects reflecting how the product becomes meaningful in one’s life.” Silverstone & Haddon’s framework of technological domestication likewise attempts to trace the evolution of a user’s relationship with a product over time (Silverstone & Haddon, 1996). Magni et al. present a model of hedonic and instrumental factors in technology exploration (Magni, Susan Taylor, & Venkatesh, 2010). And, Huang & Stolterman propose visually depicting how the functions, interfaces and goals of interactive artefacts may change over time (Huang & Stolterman, 2012). A more recent NordiCHI workshop organised to discuss the lack of “an understanding of time and the meaning of temporal aspects in UX” produced a preliminary model of users’ expectations and experiences, in the moment and as accumulated over time (Kujala, Vogel, Pohlmeyer, & Obrist, 2013).

Models of user experience tend to portray our use of technology as comprising phases of anticipation, interaction and reflection. However, it is not always clear how theories relate to design practice, how users go about evaluating their personal experiences, nor how self-reported data might enrich our understanding of our own subjective experience. In this chapter, we examine how the study of and design for experience and wellbeing might be better informed by attending more carefully to differences between reports elicited in the moment, after the fact and ahead of time. We extend previous HCI work by proposing a characterisation of experience in terms of multiple self-concepts pertinent to the study of and design for self-report. This framework, grounded in theoretical and empirical research, is presented as a heuristic and sense-making tool.

3.5 Literature Review Methodology

The framework presented in this chapter, as well as the design space described in Chapter 4, derive from a broad and interdisciplinary narrative review of the HCI, computer science and psychology liter-
Atures. In April 2015, the ACM Digital Library, PsycINFO, PsycARTICLES, JSTOR and Google Scholar databases were searched, by keyword and title where possible, using the terms Ecological Momentary Assessment, Ambulatory Assessment, Experience Sampling, Ecological Momentary Intervention, Diary Methods, EMA, EMI, and ESM. Where disambiguating terms were required, the phrases Design and Mobile Devices were appended to the search string. Search results were screened by abstract and full paper, and selected for the final analysis according to the following inclusion criteria; empirical investigations using an EMA method, user study or implementation of Ecological Momentary Assessment or Intervention (EMAI) systems, and review or analysis of the design and use of EMAI systems. This process resulted in a corpus of 286 papers, including a subset (n=42) pertaining to the relationship between EMA and conceptions of self. Further snowballing of significant references was conducted during the preparation of this chapter in early 2017, producing an additional sample (n=47) expressly related to the conception of self within HCI and psychology.

3.6 The Consolidation of Self

All experience is perceived. Some experiences are recollected, and others are not. Some shape our sense of who we are, and others do not. We refer to this complex and individual process as the consolidation of self. This chapter proposes the practical and discursive value of three self-concepts for the interpretation of experience, and in turn the design of self-report technologies; the experiencing, remembering and future-oriented selves. These characterise the consolidation of self; the perception of ‘objective experience’ by an experiencing self, its reconstruction by a remembering self, and its extrapolation by a future-oriented self perspective (See Fig 3.1).

HCI research has examined the construction and presentation of self, “a particularly salient feature of making sense of experience” (McCarthy & Wright, 2004), in role playing games (Turkle, 1994) and social networks (DiMicco & Millen, 2007).10 Models of the self in cognitive psychology are “widely
regarded to be multidimensional” (Rathbone, Conway, & Moulin, 2011), and include characterisations of past, present and future selves (Wilson, 2009), actual, ideal and ought selves, nodes in an associative memory network, hierarchical or categorical structures whose elements are traits, values or memories, multidimensional meaning spaces, and systems of self-schemas or generalisations about the self derived from past experiences (Markus & Wurf, 1987). Conway & Pleydell-Pearce’s Self Memory System (SMS) models the relationship between the self and autobiographical memory; described as comprising episodic memory, the long-term self and the working self (Conway, Singer, & Tagini, 2004). Conceptions of the self have also been studied with respect to identity “an image of the self that one tries to convey to others” (Markus & Wurf, 1987), and as extended socially, “beginning with an individual atomized self and radiating outward into the world” (Belk, 2010).

The experiencing, remembering and future-oriented selves are active, ‘experiencing’ characters which play a role in experience as lived, remembered and envisioned. Previous cognitive psychology research has also featured multiple conceptions of the self as active participants in experience, including most notably the work of Kahneman & Riis and Conner & Barrett (Kahneman & Riis, 2005; Conner & Barrett, 2012; Santangelo, Ebner-Priemer, & Trull, 2013; Zirkel, Garcia, & Murphy, 2015). These selves might also be interpreted as components of the SMS model’s ‘working’ self (Conway et al., 2004). This section grounds these selves with respect to previous research and relates their conception to the experience of self-report.

3.6.1 The Experiencing Self

“The experiencing self is, essentially, the ‘front line’ self that functions to help individuals maintain homeostasis and navigate physical and social environments in situ.” Conner & Barrett, 2012

The ancient Greeks employed two terms to describe the passage of time; ‘chronos,’ referring to the chronological passing of the hours, and ‘kairos,’ a brief, lived-in, instant of opportunity — one example of the perpetual moment in which the ‘experiencing self’ exists (Erickson, 2015).

This stage of experience is fleeting, quickly vanishing from our cognitive grasp. In the words of Kahneman & Riis, it “barely has time to exist” (Kahneman & Riis, 2005). Although this self slips from consciousness it shapes experience according to attention, mood, and other context-specific traits. The ‘experiencing self’ is continuously present yet, unintuitively, concealed from our retrospection.11

Evidence for this self is to be found in the fleeting nature of episodic memory (Conway et al., 2004) and in striking differences between momentary and retrospective reports of the same experiences (Mar-
graf, Taylor, Ehlers, Roth, & Agras, 1987; Holmberg & Holmes, 1994; Stone et al., 1998; Solhan, Trull, Jahng, & Wood, 2009; Ebner-Priemer & Trull, 2009b; Santangelo et al., 2013). Particularly pronounced effects are often reported in the context of mental health and wellbeing research. Retrospective and momentary reports of extreme mood changes, for example, have been found to possess “almost no relationship” (Solhan et al., 2009), as have reports of coping behaviours in the moment and after the fact (Stone et al., 1998). Participants in one study of panic attacks reported a fear of dying during 70% of attacks retrospectively, and in 3% in the moment. Faintness was reported during 89% of panic attacks retrospectively, and in 10% in the moment (Margraf et al., 1987). Among rheumatology patients, increasing the recall period from 1 to 7 days led to a rise in retrospective pain ratings without an associated increase in momentary reports (Broderick et al., 2008). Thirty-seven patients taking part in a study of antidepressant side-effects reported experiencing dizziness when prompted in the moment, yet only 7 reported this to their general practitioner (Barge-Schaapveld & Nicolson, 2002). These side-effects were linked to a reduction in quality of life and to treatment drop-out. These findings have significant implications for our understanding of experience, and wellbeing in particular.

3.6.2 The Remembering Self

“[T]he remembering self filters and consolidates our experiences so that we can learn, communicate, and make decisions about the future based on our past.” Conner & Barrett, 2012

The remembering self, on the other hand, is more strongly “tinged with the identity of the person” (D. J. Beal & Weiss, 2003). We are each familiar with the experience of replaying events, reliving emotions, sharing and embellishing memories — actively tending to the narrative of our lives. When we pause to reflect on past experience, we evoke this self.

The remembering self’s perspective is shaped by cognitive processes, biases and heuristics, not all of which are self-serving, and many of which we may not be aware. This includes prior beliefs (Shiffman et al., 2008), the exaggerated influence of recent and salient events (the peak-end effect) (Robbins & Kubiak, 2014), the availability heuristic, the affective valence effect, the mood congruent memory effect, duration neglect (Ebner-Priemer & Trull, 2009a, 2009b), cognitive structures, implicit theories and motivations (McFarland, Ross, & DeCourville, 1989; L. F. Barrett & Barrett, 2001; Shiffman et al., 2008), the ‘effort-after-meaning’ effect (Stone et al., 1999), the ‘telescoping’ bias (Stone et al., 1999), social desirability, and cultural norms (Scollon et al., 2009). It has been claimed that the remembering self’s perspective is as much theory as experience driven (Schwarz, 2012). As technologies mediate ever
greater percentages of our lives, the remembering self and its consolidation of experience is increasingly filtered by objects of design.

These selves have also been correlated with psychophysiological processes. Conner & Barrett show that the experiencing self is closely tied to corporeal processes and the salience network, the remembering self to memory networks. These are “different functional selves with different neurobiology” (Conner & Barrett, 2012). Cognitive psychology researchers have gone as far as to describe a ‘conflict of interest’ between these selves. The experiencing self is concerned with immediate experience, the remembering self with its memory alone (Kahneman, 2011).

3.6.3 The Future-Oriented Self

“[T]he believing self is also a ‘storyteller’ but at a higher level of abstraction that reflects the collection of identities and self concepts that help individuals maintain identity through time (2). The believing self also serves as mental scaffolding for interpreting incoming information, and for anticipating and forecasting the future, among other roles (2, 45).”

Conner & Barrett, 2012

The future-oriented, or ‘believing,’ self is engaged in expectation and aspiration, the result of “abstract, nebulous, sometimes unconscious motives” (Markus & Wurf, 1987; Rathbone et al., 2011). This self is associated with motivation, choice and individual values. Research suggests that activities of remembering and imagining overlap to the extent that it has been proposed that “the term ‘memory system’ be replaced with the term ‘remembering-imagining system,’ to “emphasise the constructive nature of remembering the past and imagining the future” (Rathbone et al., 2011; Conway et al., 2016). Mary-Helen Immordino-Yang, writing in (Calvo & Peters, 2014), states that “morally and socially complex varieties of inspiration may involve distancing oneself from the current context in order to build connections to past experiences, personal values, and possible futures.”

This self also reflects the central role of narrative in the consolidation of self; “a person’s internalized and evolving life story, integrating the reconstructed past and imagined future to provide life with some degree of unity and purpose” (McAdams & McLean, 2013). Stories are more often about “significant events and memorable moments” than “time passing” (Kahneman, 2011).
3.7 The Construal of Experience

This section examines the role these selves play in the construal of experience, each drawing on distinct sources of knowledge and engaged in a form of reflection pertinent to the design of self-report technologies. We relate these selves to the more familiar undifferentiated self, and to its own theories and concepts as most frequently encountered within HCI (See Fig 3.2).

Here, we distinguish between change to the experiencing, remembering and future-oriented selves and to the undifferentiated self, "change of a more enduring nature" (Markus & Wurf, 1987). Similar framings are offered by Markus & Wurf's dynamic self concept which defines “the self-concept of the moment” as “a continually active, shifting array of accessible self-knowledge” (Markus & Wurf, 1987), and the SMS model which describes memories as “transient constructions,” “patterns of activation over knowledge structures,” which “contain many details that are inferred, consciously and non-consciously, at the time of their construction” (Conway & Loveday, 2015). The reflection in which the experiencing self engages is dynamic — shaped by physical, social, task and temporal context (Bradley & Dunlop, 2005), attentional resources (Oulasvirta, Tamminen, Roto, & Kuorelahti, 2005) and affective states (Boehner et al., 2005). The remembering and future-oriented selves’ construal of experience is shaped by comparatively stable and consolidated features of the self, including standards, goals, attitudes (Desmet, Overbeeke, & Tax, 2001), one’s life narrative conception (Conway et al., 2004; McAdams & McLean, 2013), implicit theories (Dweck & Leggett, 1988), psychological needs (Deci & Ryan, 2000) and personality traits (Ilies & Judge, 2002). Each of these selves is engaged in a unique form of reflection.
3.7.1 Momentary Reflection

Both HCI and psychology have long featured calls for a turn towards the study of lived experience (Kjeldskov & Graham, 2003; Tamminen, Oulasvirta, Toiskallio, & Kankainen, 2003; Shiffman et al., 2008; Broens et al., 2009; X. Sun & May, 2013), the translation of laboratory findings into “real-world emotional, cognitive, or behavioral experiences” (Trull & Ebner-Priemer, 2013), and research sensitive to the progression of time (Pohlmeyer et al., 2009; Karapanos et al., 2010; Huang & Stolterman, 2012; Kujala et al., 2013).

Reviews of the HCI literature have described “a clear bias towards building systems and evaluating them only in laboratory settings, if at all” (Kjeldskov & Graham, 2003; Oulasvirta, 2009), and clinical psychology has often faced criticism for neglecting ‘the dynamics of symptoms’ (Ebner-Priemer & Trull, 2009b) given that “variability over time and dynamic patterns of reactivity to the environment are essential features of psychopathological experiences” (Myin-Germeyis et al., 2009).13

EMA, the methodology of momentary reflection, promises increased sensitivity to the passage of time (Stone et al., 1999), enhanced ecological validity,14 and a more accurate causal understanding of events, allowing antecedent conditions, moderating variables and outcome measures to be delineated “much more effectively than panel or retrospective designs,” as well as permitting designers to better understand “users’ underlying motivations” (Pielot, Poppinga, & Boll, 2009) and potentially “unearth unexpected behaviors and adaptations” (Carter, Mankoff, & Heer, 2007).15

The ‘white coat hypertension effect’ is a classical example of the importance of ecological validity. Measures of blood pressure conducted in clinical settings consistently produce readings higher than those found in daily life (Smyth & Stone, 2003; Ebner-Priemer & Trull, 2009b; Wilhelm & Perez, 2013; Robbins & Kubiak, 2014). A similar effect has been found for endocrine reactivity (Smyth & Stone, 2003), implying that “hundreds of thousands of people may have been misdiagnosed, and possibly mistreated” (Ebner-Priemer & Trull, 2009b). The diagnostic criteria for many mental health disorders hinge upon characteristic cycles of behaviour. Individuals with bipolar disorder, for example, are believed to demonstrate significantly greater affective instability than those with depression (Trull et al., 2008). One study conducted among college students (n=108) found that the variability, rather than intensity, of depressive affect predicted suicidal ideation and suicide attempts (Witte, Fitzpatrick, Joiner, & Schmidt, 2005). Another (n=36) found that EMA measures of affect predicted non-suicidal self-injurious behaviour (Armey, Crowther, & Miller, 2011).16

Although the adoption of EMA is often motivated by a need to support ecological validity, mo-
mentary reflection is itself susceptible to bias, as a result of reactivity effects, habituation to reporting, and psychological defence strategies such as the maintenance of social desirability (Barta, Tennen, & Litt, 2012). Measuring and mitigating these biases is essential to the pursuit of a valid understanding of momentary experience (Scollon et al., 2009).

Slipping from our cognitive grasp, the experiencing self embodies a potential for knowledge — an opportunity to recapture this fleeting glimpse of experience in time.

3.7.2 Retrospective Reflection

Most research pertaining to the study of human behaviour employs retrospective forms of reflection (Brinkman, 2009). Much of our accumulated knowledge of human experience therefore rests on an assumption of accurate recall. Clinical psychology, for example, has faced criticism for the fact that “retrospective self-reports of patients’ symptoms,” including clinical interviews and screening scales, remain the primary, and often sole, source of practitioner knowledge (Ebner-Priemer & Trull, 2009a). We might likewise bring to mind the percentage of design research which asks users to recall their experience of a device, website or mobile application.

Interviews and questionnaires offer relative ease of application and can provide unique insight into conscious processes (Brinkman, 2009). Their weaknesses are also well-documented however, including their potential to constrain participants’ responses, and sensitivity to anchoring and priming effects (Schuman & Presser, 1979; Schwarz, 2012). As Kahneman & Riis write, “It is a common assumption of everyday conversation that people can provide accurate answers to questions about their feelings, both past (e.g. ‘How was your vacation?’) and current (e.g. ‘Does this hurt?’)" (Kahneman & Riis, 2005). However, numerous studies have demonstrated that the ‘remembering self’ frequently violates conventional logic (Margraf et al., 1987; Holmberg & Holmes, 1994; Stone et al., 1998; Solhan et al., 2009). When a series of couples were asked to evaluate their honeymoon both at the time and years later, those who had experienced relationship problems during the intervening period rated the experience more negatively than they had at the time (Holmberg & Holmes, 1994). Another EMA study (n=828) found that teenagers who spent more time studying were happier despite the finding that studying itself evoked lower happiness than most other activities (Csikszentmihalyi & Hunter, 2003). Low levels of correlation have also often been reported between subjective and objective measures of physical activity (Ebner-Priemer & Trull, 2009b; Santangelo et al., 2013).

The remembering self’s conception of experience is not fixed but revisited, reconstructed, and can lead to divergent reports of experience with the progression of time. Often it is not how well a user...
remembers an experience which best supports the action of design but why certain details, rather than others, persist upon reflection (Bluck, Alea, Habermas, & Rubin, 2005).

### 3.7.3 Prospective Reflection

Temporal distance changes the way we mentally represent experience, and shapes our attitude towards future events. Events further into the future are more likely to be represented by a small number of abstract features which convey their “perceived essence ...(high-level construals),” rather than “more concrete and incidental details ...(low-level construals)” for example (Trope & Liberman, 2003). Distant-future preferences are therefore more likely to be influenced by cognitive goals and perceptions of desirability, near-future preferences by affective goals and perceptions of feasibility (ibid).

Although the fallibility of prospective reflection is more easily grasped, its scale can still prove surprising. The Service Recovery Paradox serves as an example of ostensibly contradictory choice. Customers often report greater satisfaction with a service following recovery from failure than with respect to one which has never failed (De Matos, Henrique, & Rossi, 2007). In an experiment often cited as a quintessential example of duration neglect and the peak-end effect, 80% of participants, when asked to repeat one of two experiences, chose to immerse a hand in painfully cold water for a longer period of time when the water’s temperature was slightly reduced for the final 30 seconds — “apparently preferring more pain over less” (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993). Recent HCI studies have found peak-end effects with respect to screen-based interaction (Cockburn, Quinn, & Gutwin, 2015) and computer gaming (Gutwin, Rooke, Cockburn, Mandryk, & Lafreniere, 2016).

The tight circle of human-computer interaction often demands rapid decision-making, repetition and prolonged engagement. The design choices which best support these activities are most effectively informed by very different criteria if the desire to repeat an experience is influenced more strongly by its peak, or end, than duration.

### 3.7.4 Multiple Consciousness

These modes of reflection are not necessarily, and perhaps rarely, mutually exclusive but combined in a state of ‘multiple consciousness.’

“We are involved in the experience, but even as we are involved we are already witnesses with one eye to the past and another to the future. In a sense, we live an experience in the light of how we are likely to recount it to others.” McCarthy & Wright, 2004

The experiencing, remembering and future-oriented selves are dynamic entities, “active, forceful,
and capable of change” (Markus & Wurf, 1987). Reflection is a feature of daily life; at times invoked with intent, as during research, and often encouraged by design. Our use of technology to share experience, by means of text-messaging, social media and selfie-taking for example, often evokes particular states of multiple consciousness (Grinter, Palen, & Eldridge, 2006; Cui & Wang, 2012; Souza et al., 2015). These activities allow us to record our experience while exercising control over our self-image and expressing an idealised narrative identity. These self-portrayals reach large audiences, are intrinsically meaningful, and may heighten the reflection we engage in. Design therefore shapes how we construe experience, and present ourselves to others. The wording of status requests, the structure of social networks, the delivery of notifications, the timing of interaction, and the aesthetic presentation of data all entail such effects.

This framework of multiple selves provides a scaffolding for the construal of experience, the practice of self-report, and the design and use of the technologies through which we live, remember and envision our lives.

3.8 Conflicting Self-Reports

We have seen that self-reports of the same experience do not always match. These disparities are often viewed as a form of error but are also an opportunity to learn more about users and experience. We next introduce a new concept, inter-self dissonance, to facilitate such interpretation.

3.8.1 Ground Truth & Self Truths

Inter-self dissonance refers to the coexistence of multiple, incongruent perspectives on experience, each realised through experiential construal (See Fig 3.3). Instead of interpreting these disparities purely as the result of erroneous representations of experience however, we might consider each of the experi-
encing, remembering and future-oriented selves to yield a unique form of ‘self-truth.’

This concept of multiple truths has important implications for the design and analysis of self-report technologies, as for the lived interpretation of experience. These incongruities are largely imperceptible to the selves which comprise their existence. The fact that the experiencing self, for example, persistently evades our cognitive grasp, is likely to impede awareness of dissonance pertaining to momentary experience.

3.8.2 Self Knowledge \([d_{OE}, d_{ER}, d_{OR}]\)

HCI is increasingly focused on the design of technologies to facilitate self-report, not always for the purposes of research but to support self-knowledge, behaviour change and wellbeing. Evidence of stark differences between temporally-distinct reports of experience compels us to ask what effect revealing these incongruities to users might have. Might this prove insightful, persuasive, or support a richer understanding of subjective experience? One study described in Section 3.6.1 recounted remarkable variation between momentary and retrospective reports of panic attacks (Margraf et al., 1987). We might hypothesise that highlighting such discrepancies to users could yield a therapeutic effect, supporting a healthier and more coherent construal of experience for example. In fact, De Beurs et al. studied 32 patients’ experience of panic attacks over a 12 week period (De Beurs, Lange, & Van Dyck, 1992). Before treatment, participants significantly overrated panic attacks reported retrospectively, compared to continuous monitoring. After treatment, both methods ‘matched perfectly.’\(^{18}\)

Research suggests that those who “exhibit discrepancies between implicit and explicit measures of their self-concepts or motives” may be “especially low in emotional well-being and especially high in physiological reactivity, anxiety, self-doubt, defensiveness, and narcissism” (Wilson, 2009).\(^ {19}\) In an unpublished study conducted by Kubiak et al., and described in (Robbins & Kubiak, 2014), an electronic diary was employed to improve awareness of hypoglycemia among participants with type 1 diabetes. Following completion of a timed test of alertness and motor speed, participants were “prompted to estimate their performance and received immediate feedback on their actual performance,” in line with the authors’ underlying rationale to “help the patient notice subtle decrements in cognitive performance … common early signs of falling blood glucose levels.”\(^ {20}\)

These examples suggest the capacity of inter-self dissonance to support a richer understanding of subjective experience from self-reported data.
3.8.3 Future Choice \([d_{RF}, d_{EF}, d_{OF}]\)

“description and prediction are different goals and their optimization requires different strategies.” Schwarz, 2012

Inter-self dissonance can also result in poor decision-making. If the remembering self holds a skewed impression of past experience then, although “biased estimates of the past can be good predictors of future choice” (Conner & Barrett, 2012), this may not lead to ‘good’ choices. Essentially, the reconstruction of experience impairs learning (Schwarz, 2012). This applies not only to individuals but also to research and professional practice, including clinical psychology (Ebner-Priemer & Trull, 2009a).

While the remembering and future-oriented selves may not draw on an accurate portrayal of experience, it is this depiction which best predicts an individual’s future actions. This statement has implications for design. In the analysis of behaviour change and decision-making, for example, self truths can prove more informative than ground truth (Hicks et al., 2010). During one study, recollected, and not experienced, enjoyment predicted participants’ desire to repeat a vacation (Wirtz, Kruger, Scollon, & Diener, 2003). The term ‘miswanting’ was coined to describe the role of affective forecasting with respect to poor decision-making (See Section 2.6.2) (Gilbert & Wilson, 2000). When asked how much they enjoyed using a system, a user will often provide an answer to a subtly yet notably different question; how much they enjoyed that experience, when they pause to reflect upon it.

Some design researchers have questioned the advantage of momentary reports, arguing, for example, that “it may not matter how a product was experienced in a given situation, but what individuals remember from this experience” (Karapanos et al., 2010). However, our interpretation of experience is shaped by how we both live and reflect upon it (Kahneman & Riis, 2005). Design is often driven by our understanding of users’ choices; whether, and why, to use, disengage from or return to a particular system, for example. These aims benefit from an understanding of the weight of multiple forms of reflection with respect to decision-making.

3.8.4 The Drive to Reduce Dissonance

Research suggests that we are driven to reduce dissonance between activities of anticipation, experience and recollection. Mitchell et al. describe several such tendencies which can be aligned with a framework of inter-self dissonance; ‘rosy prospection,’ a tendency to view future events more favourably than they will be experienced, ‘rosy retrospection,’ a tendency to remember past events more favourably than
experienced, and a consistency effect, a tendency to align the memory of an event with prior expectations (Mitchell, Thompson, Peterson, & Cronk, 1997). Experiments have shown that these phenomena are “associated with an increase in the number of negative thoughts during the event which seem to be caused by distractions, disappointment, and a less positive view of the self” (Mitchell et al., 1997). We continuously reinterpret our experiences, which, compared to material possessions, “make people happier,” precisely because they “are more open to positive reinterpretations, are a more meaningful part of one’s identity, and contribute more to successful social relationships” (Van Boven & Gilovich, 2003). As van Boven & Gilovich state, we enjoy the anticipation of experience and further ‘consume’ its recollection (ibid.).

3.9 Designing for Multiple Selves

This chapter is oriented not towards the prescription of an overarching cognitive model of self but the practical and discursive value of an epistemological perspective of multiple self-concepts of a sufficiently simple form to function as a practical heuristic and sense-making tool.21

3.9.1 The Importance of Multiple Time-Frames

The outcomes we aim to support through design, including wellbeing, often consist of components “at least partly independent of each other” with respect to their evaluation (Kahneman & Riis, 2005). Studies have shown that our knowledge of wellbeing hinges upon how it is both lived and reflected upon, that “positive emotions belong to experienced utility; meaning and life satisfaction …to evaluation.” As Kahneman & Riis write however, “the dominant practice in well-being research effectively ignores the issue, and continues to treat the determinants and consequences of happiness as if it were a unitary concept” (Kahneman & Riis, 2005). Although HCI researchers have stressed the need to evaluate systems ‘in the wild’ (Kjeldskov & Graham, 2003; Consolvo et al., 2008; Rennick-Egglestone et al., 2016), the implications of multiple modes of self-report for design remain under-explored.22 This framework heightens our attention to the nature of experience as lived, remembered and envisioned.

3.9.2 Drawing on Self-Knowledge

Sellen & Whittaker write that “[r]ather than try to capture everything, system design should focus on the psychological basis of human memory” (Sellen & Whittaker, 2010). This requires a sensitivity not only to what but how we think about experience. This heuristic of multiple selves provides a scaffolding
for research and design, highlighting how the practice of self-report taps divergent sources of knowledge, no single time-frame defines experience, and experience is conceived of multiple consciousness.

3.9.3 A Richer Understanding of Experience through Self-Report

The subject matter of many technologies is experience; its report, reflection and exploration. Mobile applications for mood, exercise and sleep tracking, for example, often strive to help users make sense of their own experiences (Hekler, Klasnja, Froehlich, & Buman, 2013; Hollis, Konrad, & Whittaker, 2015). Key design challenges for these systems include the engagement of users and the extraction of actionable insights from data. A common criticism these technologies face however, is that, although they facilitate tracking, the insights they provide rarely extend beyond drawing the blinds, closing the laptop and cutting down on that last cup of coffee before bed. In the concept of inter-self dissonance, we find a potential catalyst for change, driven by awareness, in the moment, of the disparity between one’s past, present and possible-future experiences of bed-time.

3.9.4 Tools for Reflection

Much HCI research pertains to how we think about experience as much as to interaction itself. Many researchers are concerned with the design of technologies to support reflection (Hallnäs & Redström, 2001; Peesapati et al., 2010; Isaacs et al., 2013; Hollis et al., 2015), learning (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012), behaviour change (Hekler et al., 2013; Siegel & Beck, 2014; Hollis et al., 2015), wellbeing (Thieme, Wallace, Meyer, & Olivier, 2015), the presentation of self (DiMicco & Millen, 2007; Dong, Qiu, & Zhu, 2014) and the recording of experience (Sellen & Whittaker, 2010). These technologies do not so much instruct users as engage us in conversation, as participants in the creation of experience, posing questions as often as addressing them.

Health and wellbeing researchers have begun to recognise a “need for users to be able to reflect on their experience” (Rennick-Egglestone et al., 2016). Many technologies already serve as observational tools, supporting our capacity to turn our attention inwards. Reflection on past experience is not just a behaviour but a skill however, a distinction exemplified in the development of MAHI, an application designed to help diabetes patients acquire reflective thinking skills (Mamykina, Mynatt, Davidson, & Greenblatt, 2008). MAHI employs breakdowns in understanding, characterised by a “fleeting sense of meaning,” as a “trigger for reflection” (Mamykina et al., 2008) — a framing which shifts the focus of design from activating to enriching reflection.

Design for memory and remembering is another emerging thread of HCI research which presents
unique challenges. More often than not, these are not “activities people set out to do,” and, as the time we spend with digital technologies increases, the remembering in which we engage is more likely to be externally cued (Norman, 2009; van den Hoven, 2014; Van Gennip, Van Den Hoven, & Markopoulos, 2015). A nascent vocabulary of ‘technology-mediated reflection’ (Isaacs et al., 2013) and ‘remembering experiences’ (van den Hoven, 2014) points to a design practice already grappling with multiple modes of reflection.

3.9.5 Evoking the Selves in Design

One approach to design in light of a heuristic of multiple selves is to consider ‘design to evoke a specific self.’ A framing of multiple selves is compatible with many design methods, including personas, scenarios and storyboards. Harper et al. have studied the ‘temporal experiencing’ of Facebook and propose that its features lead “people to feel as if they are always acting ‘in the now,’” and cause their history “to disappear from view” (Harper, Whitworth, & Page, 2012). They describe this as a ‘fixing’ of the self, writing that “when people post a status update on Facebook, or, by the same token, when they read and comment on such a comment produced by others, they are not simply living in the durée but organising themselves, defining themselves, and the experience of themselves vis-à-vis their relations with others” (Harper et al., 2012). An application for mood-reporting encourages users to “associate a bad habit with long-term emotional consequences, rather than the generally positive association of immediate relief” (Hollis et al., 2015). Another attempts to “sub-consciously or semi-consciously notify the gamer of how long they have been sedentary” by dimming the screen (Garg, Adams, Husted, & Brown, 2014). This is design to evoke a particular self.

3.10 Conclusion | Reflections & Takeaways

The framework of multiple selves presented within this chapter heightens our attention to the importance of how we both live and think about experience. Differences between reports of experience made in the moment, after the fact and ahead of time have significant implications for our understanding of wellbeing, the design of self-report technologies, and the practice of mental health care. A sensitivity to multiple time-frames therefore enables a more accurate portrayal of the report, interpretation and curation of experience.

This chapter persistently articulates the need to refrain from approaching wellbeing uncritically — as fixed, well-understood or unrelated to the experience of self-report. This attention to our orientation in time is maintained throughout this work, ensuring, for example, salient discussion of experience
and reflection during design research conducted with women and health professionals, motivating the inclusion and informing the design of both momentary and retrospective reporting mechanisms during the development of a protocol and mobile application for the self-report of wellbeing during pregnancy, and providing a coherent framing for the future analysis of a corpus of psychological wellbeing with respect to multiple time frames.
Chapter 4

The Design of Self-Report Technologies

The average delay between the onset of mental illness and its treatment is 11 years (Wang, Berglund, Olsson, & Kessler, 2004). At any one time, less than 1 in 3 of those meeting diagnostic criteria for depression are in receipt of care (Ormel et al., 2008). The challenges of mental healthcare provision have been described as “so long standing, so vast, and so unresponsive” that they require a new approach entailing ‘levers of change’ on a population scale (Atkins & Frazier, 2011). This is the promise of mobile technology; insight into the lived experience of mental health and support for those who might not otherwise seek help (Shiffman et al., 2008; Atkins & Frazier, 2011; Wichers et al., 2011).

4.1 Technology’s Reach

“[M]obile technology is poised to become the most powerful form of media to influence clinical practice.” Rizvi, Dimeff, Skutch, Carroll, & Linehan, 2011

These are problems of reach, access and disclosure. Mobile applications have allowed researchers to engage greater numbers of participants in cognitive psychology (n=4,157) (Dufau et al., 2011), mind-wandering (n=5,000) (Killingsworth & Gilbert, 2010) and happiness (n=45,000) (Miller, 2012) research. Mobile technologies might also allow public health services to extend their reach; physically, socially and psychologically.

The use of these devices in public health screening could increase the number of patients screened while reducing costs and the administrative burden associated with paper questionnaires. This would in turn increase access to treatment and enable a more proactive approach to care (Wichers et al., 2011).
The intimacy which many of us associate with our personal devices has enabled researchers to study actions and events which “take place ‘behind closed doors’ (literally and figuratively)” (Zirkel et al., 2015), including how children and teenagers spend their time (Larson & Richards, 1991; Strack, Magill, & McDonagh, 2004), children’s and parents’ experiences of Attention Deficit Hyperactivity Disorder (ADHD) (Whalen et al., 2006), the social reality of young families (Aharony, Pan, Ip, Khayal, & Pentland, 2011), everyday sexism (Swim, Hyers, Cohen, & Ferguson, 2001), racism (Swim, Hyers, Cohen, Fitzgerald, & Bylsma, 2003) and stress (Perrez & Reicherts, 1996).

Mental health screening is further complicated by the need to support disclosure while overcoming a pervasive mental health-related stigma (Goodman, 2009; Howard et al., 2014; Kingston et al., 2015). Despite heightened attention to wellbeing during pregnancy, for example, concerns often go unreported as a result of these barriers to care (Moore et al., 2016). Mobile devices might also support the engagement of at-risk populations, including those with mental illness. Participants with schizophrenia, bipolar disorder and substance dependence have shown “good compliance and low dropout rates in general” during previous research (Freedman, Lester, McNamara, Milby, & Schumacher, 2006; Granholm, Loh, & Swendsen, 2008; Trull & Ebner-Priemer, 2013).

4.2 Fulfilling the Promise of EMA

Mobile devices have the potential to transform public health, clinical psychology, human-centred research, and design practice (Shiffman et al., 2008; Lazer et al., 2009; Dufau et al., 2011). However, significant barriers to their use and adoption remain. As Shoham & Insel write, complaints are rarely “about the dearth of technology” but “about its excesses, turning patients into ‘i-patients’ ” (Shoham & Insel, 2011). This chapter charts the design space for self-report technologies with an eye towards supporting the adoption of EMA systems for research and clinical practice, as pertains to challenges of user engagement, reporting burden, data validity and honest disclosure among others (Fahrenberg, Hüttner, & Leonhart, 2001; Smyth & Stone, 2003; Ebner-Priemer & Trull, 2009b; V.-J. Khan, Markopoulos, & Eggen, 2009; Miller, 2012).

4.3 The EMA Technology Design Space

The feasibility of EMA hinges in large part upon technology. This section examines the characteristics of effective EMA systems, drawing on a review of those described and deployed within the HCI and psychology literatures (See Tables 4.1 & 4.2) as well as available commercially.
### Table 4.1: EMA Technologies [Pre-2005]

<table>
<thead>
<tr>
<th>Title</th>
<th>First Date Referenced</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMRES</td>
<td>1987</td>
</tr>
<tr>
<td>(Perrez &amp; Bechert, 1987; 1996; Bolger, Davis, &amp; Rafaeli, 2003)</td>
<td></td>
</tr>
<tr>
<td>Electronic Diary</td>
<td>1992</td>
</tr>
<tr>
<td>(Perren, Shiffman, Paly, &amp; Fritzsche, 1994; Stone et al., 1998; Bolger et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>PMC-diary</td>
<td>1999</td>
</tr>
<tr>
<td>(Bolger et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>The Experience Sampling Program (ESP)</td>
<td>1999</td>
</tr>
<tr>
<td>(Bolger et al., 2003; D. J. Barnett &amp; Berntz, 2005; Le, Chia, &amp; Brat, 2006; V. Khan &amp; Markopoulos, 2009; Fischer, 2009)</td>
<td></td>
</tr>
<tr>
<td>FASEM-C</td>
<td>2000</td>
</tr>
<tr>
<td>(Perrez, Schooler, &amp; Wilhelm, 2000; Wilhelm &amp; Perrez, 2013)</td>
<td></td>
</tr>
<tr>
<td>MONITOR</td>
<td>2001</td>
</tr>
<tr>
<td>(Fahrenberg et al., 2001; Eitem-Premier &amp; Kubalek, 2007)</td>
<td></td>
</tr>
<tr>
<td>ISIS for Pocket PC</td>
<td>2001</td>
</tr>
<tr>
<td>(Bolger et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>The Electronically Activated Recorder (EAR)</td>
<td>2001</td>
</tr>
<tr>
<td>(PMI, Prenatal Care, Child, &amp; Price, 2002; Bubion &amp; Kubalek, 2014)</td>
<td></td>
</tr>
<tr>
<td>PsyctLab Mobile (evolved from EAR)</td>
<td>2001</td>
</tr>
<tr>
<td>(Eitem-Premier &amp; Kubalek, 2007; V. Khan &amp; Markopoulos, 2009)</td>
<td></td>
</tr>
<tr>
<td>EntryWare</td>
<td>2002</td>
</tr>
<tr>
<td>(Gavrilov, 2002; V. Khan &amp; Markopoulos, 2009)</td>
<td></td>
</tr>
<tr>
<td>CAES</td>
<td>2003</td>
</tr>
<tr>
<td>(Intille, Rondoni, Kubalek, Ancora, &amp; Bas, 2001; V. Khan &amp; Markopoulos, 2009)</td>
<td></td>
</tr>
<tr>
<td>ChronoRecord</td>
<td>2003</td>
</tr>
<tr>
<td>(Whiteside et al., 2003; M. Bauer et al., 2004)</td>
<td></td>
</tr>
<tr>
<td>iESP (evolved from ESP)</td>
<td>2003</td>
</tr>
<tr>
<td>(Comiskey &amp; Walker, 2003; Le et al. 2006; Fischer, 2009)</td>
<td></td>
</tr>
<tr>
<td>Transactional Electronic Diary (TED)</td>
<td>2004</td>
</tr>
<tr>
<td>(Gohalszadeh, Epstein, Moghershahi, Lin, &amp; Preston, 2004)</td>
<td></td>
</tr>
<tr>
<td>PMAT</td>
<td>2004</td>
</tr>
<tr>
<td>(Wells, Best, Lucy, &amp; MacDermid, 2004; V. Khan &amp; Markopoulos, 2009)</td>
<td></td>
</tr>
<tr>
<td>ContextPhone</td>
<td>2005</td>
</tr>
<tr>
<td>(Ranter, Oudshoorn, Piet, &amp; Stoemen, 2005; Miller, 2012)</td>
<td></td>
</tr>
<tr>
<td>SocioXensor</td>
<td>2005</td>
</tr>
<tr>
<td>(Mulder, Ter Hofte, &amp; Kurt, 2005; Ter Hofte, Ott, Redemors, &amp; Mulder, 2006; Ter Hofte, 2007; Miller, 2012)</td>
<td></td>
</tr>
</tbody>
</table>

#### 4.3.1 Modes of Use

Design entails choice in the service of use. In the case of self-report technologies, choice often pertains to questions of engagement, burden, validity, change and disclosure; values entangled in tensions of scale, intrinsic and extrinsic motivation, nomothetic and idiographic use, individual, normative and clinical outcomes, self and other.9

The design values expressed by developers of self-report technologies often include efficiency, ease of use, minimal device performance impact, minimal disruption, data privacy, security, wellbeing, engagement, minimal burden, satisfaction and data quality (Froehlich et al., 2007; Hicks et al., 2010). Enjoyment, empowerment, learning, responsibility, transparency, transformation, autonomy, competence, sociality, belonging, meaning and self-esteem tend to receive a less explicit focus.10

An application for ADHD screening among teenagers which shares users’ data with a healthcare professional is likely to result in very different reporting practices than a similar application designed to support personal self-awareness. Users may be motivated to avoid judgement or enable access to support services.11 These practices are shaped by design. Although introducing a mindfulness exercise
Table 4.2: EMA Technologies (Post-2005)

<table>
<thead>
<tr>
<th>Title</th>
<th>First Date Referenced</th>
</tr>
</thead>
<tbody>
<tr>
<td>MyExperience</td>
<td>2007</td>
</tr>
<tr>
<td>Momento</td>
<td>2007</td>
</tr>
<tr>
<td>Reconexp</td>
<td>2008</td>
</tr>
<tr>
<td>EpiCollect</td>
<td>2009</td>
</tr>
<tr>
<td>Mood24/7</td>
<td>2011</td>
</tr>
<tr>
<td>DBT Coach</td>
<td>2011</td>
</tr>
<tr>
<td>Funf</td>
<td>2011</td>
</tr>
<tr>
<td>SERENA</td>
<td>2013</td>
</tr>
<tr>
<td>Emotion Sense (also Easy M and Q Sense)</td>
<td>2013</td>
</tr>
<tr>
<td>PIEL Survey</td>
<td>2014</td>
</tr>
<tr>
<td>LocasaESM</td>
<td>2014</td>
</tr>
<tr>
<td>Purple Robot</td>
<td>2014</td>
</tr>
<tr>
<td>PACO</td>
<td>2014</td>
</tr>
<tr>
<td>Gquest</td>
<td>2014</td>
</tr>
<tr>
<td>Aware</td>
<td>2015</td>
</tr>
<tr>
<td>SurveySignal</td>
<td>2015</td>
</tr>
</tbody>
</table>

This chapter explores the dials at the hands of the EMA technology designer.

4.3.2 EMA Technology Features

The first momentary assessments were completed on paper. Timed bleepers, Personal Digital Assistants (PDAs) and smartphones each extended the reach, usability, validity and reliability of these methods. Even today, a large percentage of the EMA literature refers to the design of PDA software (Stone & Shiffman, 2002; Christensen et al., 2003; Le et al., 2006; Ebner-Priemer & Kubiak, 2007).12

In 2017, 94% of UK adults report owning a mobile phone, and 76% a smartphone (Ofcom, 2017). The flexibility, interactivity, intimacy and connectivity of these devices has greatly facilitated the self-report of lived experience.

The capacity of mobile devices to capture ancillary data including timestamps has proved essential
to the feasibility of EMA. Using paper scales, participants were often found to engage in ‘parking lot compliance,’ completing reports only moments prior to their submission (Smyth & Stone, 2003). One study found an actual compliance rate of 11% for paper diaries compared to a reported rate of 90% (n=80) (Stone, Shiffman, Schwartz, Broderick, & Hufford, 2002). Hormuth wrote in 1986 that questions concerning reliability and validity lacked answers; “Does the subject respond to the signals on time? Is the subject’s objective circumstances influenced by participation? Is the subject’s subjective perception of a situation influenced by the method? Is the subject capable of reporting and rating situations?” (Hormuth, 1986).

Mobile devices have allowed researchers to approach answers to many of these questions while reducing the burden on users, through the automated collection of interaction logs, geopositioning, accelerometry and connectivity data. Effective EMA technologies permit researchers and practitioners to register participants, send alerts, present questions, accept responses, provide feedback, and debrief users. How scales are presented to users matters. The use of frequency scales, text boxes of different sizes, question layouts and scrolling have all been found to affect validity and engagement (Mavletova & Couper, 2014; Lanzola et al., 2014; Wells, Bailey, & Link, 2014). Several researchers have also explored unique visual representations of scales, including the paper-based interpersonal grid (Fournier, Moskowitz, & Zuroff, 2008) and Inclusion of Other in the Self (IOS) scale (Aron, Aron, & Smollan, 1992), interactive digital presentation of the circumplex model of emotion (MoodMap) (Morris et al., 2010), and a minimal stress scale comprising a single question (AndWellness) (Hicks et al., 2010). The creative potential of mobile devices to elicit and express users’ inner experiences remains largely under-explored however.

Reporting guidelines for momentary assessment stress the “mountain of data” which these techniques can produce (Stone & Shiffman, 2002). A further advantage of EMA technologies is therefore their capacity to facilitate data management, transfer, analysis and modelling. The mobile app AndWellness, for example, provides a survey authoring tool and a visualisation toolkit (Hicks et al., 2010). The core functionality of the extensible system Purple is described in terms of user management, content authorship, content delivery and data management capabilities (Schueller et al., 2014). Other applications have implemented additional features for the purposes of user engagement, reflection, analysis and data sharing, including information-provision, goal-setting, note-taking and scheduling tools, visualisations, interventions, and sharing functionalities (Aharony et al., 2011). MoodMap, an application designed to support emotional self-awareness, for example, provides cognitive reappraisal, physical relaxation and breathing visualisation exercises in addition to facilitating mood tracking (Morris et al., 2010).
4.4 Designing EMA Protocols

In general, optimal EMA protocols have yet to be developed (Shiffman et al., 2008). Methodological and technological considerations are closely interlinked however, and this section provides insight into their relationship and design. Researchers have variously described interval, signal and event-contingent (Bolger et al., 2003; Christensen et al., 2003), schedule, frequency and timing based (Shiffman et al., 2008), and event and time based (Shiffman et al., 2008) protocols.

4.4.1 Time-Contingent Sampling

Many EMA protocols are time-contingent. Users are notified to provide reports according to a semi-random or fixed schedule. A typical approach employs a coverage strategy — semi-random sampling within fixed intervals. A user might be prompted randomly within every 3 hour period, for example. This approach aims to minimise adaptation to a predictable schedule (D. J. Beal & Weiss, 2003). The choice of reporting frequency is typically context-dependent (Ebner-Priemer & Trull, 2009b). While retrospective, global or trait measures can prove ‘too cold,’ “too slow and sluggish to change,” momentary measures can also be ‘too hot,’ “too volatile and overly sensitive to extraneous variables” (Conner & Barrett, 2012). However, there currently exists “a paucity of studies comparing various time-based designs” and there are “no general conventions,” a fact which Santangelo et al. write is unsurprising, “as the temporal dynamics of emotional and cognitive processes are largely unknown” (Santangelo et al., 2013).

Many studies have employed intensive sampling protocols. Nine hundred and ninety-one Scottish adolescents were prompted to report their current activity every 15 minutes outside of school hours for 4 days (Biddle, Gorely, Marshall, & Cameron, 2009). One hundred and twenty-four users were signalled to report instances of mind-wandering 8 times per day for 7 days (Kane et al., 2007). And in another example, almost 200 participants were asked to complete assessments 5 times per day for 1 week every 5 years as part of a longitudinal study of emotional wellbeing (Carstensen et al., 2011). Other researchers have advised that sampling more than 6 times per day for periods longer than 3 weeks should be avoided unless assessments are particularly short or additional incentives are provided (See Delespaul, 1992 as cited in Christensen et al., 2003).

There is a need to tailor protocol design to the phenomena of interest. Consider the case of an EMA application designed to screen for anxiety by means of the repeated assessment of worry in daily life. How often, or under what circumstances, should worry be evaluated? We are required to make
an initial characterisation of the phenomena of interest, possible contingent variables, and appropriate reporting mechanisms.

“A central consideration in ESM research is that the research itself really must be theory driven. Although theory is valuable in the design of any study, ESM demands that the research make choices that only theory can guide.” Zirkel et al., 2015

In practice, EMA protocols are often built around context-specific design hypotheses. In this example, this might concern the rate at which worry is likely to fluctuate, shaped by attempts to reduce the bounds of error (such as more frequent sampling) and the burden on users (such as less frequent sampling).

4.4.2 Event-Contingent Sampling

Event-contingent sampling is more often employed when targeting rare or highly specific events. This requires the choice of a ‘trigger’ (Lathia, Rachuri, Mascolo, & Rentfrow, 2013). Users may be prompted to provide reports according to an ancillary measure of context, such as a change of location, or asked to complete assessments according to predefined subjective criteria, such as eating a meal or exercise. Five hundred and sixty-two adolescents participating in one study of smoking behaviours were trained to report ‘smoke’ and ‘no smoke’ events in addition to responding to random prompts 5 or 6 times a day for 4 periods of 1 week, 6 months apart, for example (Mermelstein, Hedeker, Flay, & Shiffman, 2007).18 Sampling protocols might also adapt to changes in users’ behaviour over time (Lathia, Rachuri, Mascolo, & Rentfrow, 2013). One system tailored prompting to users’ sleep cycles, according to the times at which morning and evening diaries were completed (Sorbi, Mak, Houtveen, Kleiboer, & van Doornen, 2007). The use of objective triggers eliminates reliance on a user’s interpretation of what constitutes an event, and enables the assessment of compliance (D. J. Beal & Weiss, 2003). However, signal-contingent protocols can mean that participants are more likely to be prompted at inopportune moments. One study found that those providing reports according to subjective criteria forgot to do so, and those signalled objectively “deliberately did not answer because it was too much effort” (Möller et al., 2013).

4.4.3 Sampling Duration

The length of time for which sampling is conducted is typically dependent upon the time-frame most “likely to reveal dynamic processes …of interest” or the duration of the period during which “change is likely to occur” (Bolger et al., 2003). Burst protocols, entailing contiguous sampling periods separated
by extended periods of time, from days to years, have occasionally been employed to support longitudinal reporting (Carstensen et al., 2011). In such instances, it may be necessary to account for routine and weekend effects, by commencing assessment periods on different days for example (Fahrenberg et al., 2001).

Theories of behaviour change can also inform sampling protocol design (Schüz, Cianchi, Shiffman, & Ferguson, 2015). These theories provide insight into the micro-processes of behaviour and may enable assessment based on “an understanding of how affect, cognition, and behavior interact and unfold over time” (Shiffman et al., 2008). The Transtheoretical Model of Behavioural Change, for example, posits ten processes implicated in progress through six stages of change (precontemplation, contemplation, preparation, action, maintenance and termination) and has been applied to a wide variety of problem behaviours including smoking cessation, cocaine addiction, weight control, high-fat diets, adolescent delinquent behaviours, safer sex, condom use, sunscreen use, radon gas exposure, exercise acquisition, mammography screening and physicians’ interaction with smokers (Velicer et al., 1993; Prochaska et al., 1994; Prochaska & Velicer, 1997).

The challenge of imposing an acceptable burden on users is defined by compromise — between the number of reports requested and the time required to complete assessments in particular (Scollon et al., 2009). Some researchers have advised that momentary reports should take no longer than 2 minutes to complete (Hormuth, 1986; Christensen et al., 2003; Consolvo & Walker, 2003). Although, older studies often reported completion times of 5 minutes or longer (Perrez & Reicherts, 1996). Three hundred and eleven participants in one study of stress and coping were prompted to provide reports 6 times per day for 1 week, and took between 4.65 and 7.28 minutes to complete assessments on average. Seventy-six percent of users found the reporting duration acceptable and 94% judged the experience positively (Perrez et al., 2000).

‘Liveability functions,’ including silent, snooze or do-not-disturb modes and repeated prompting strategies, have also been employed as a means to support engagement (Santangelo et al., 2013). Morris et al. allowed participants to choose prompting intervals between 30 minutes and 3 hours in duration, although not the exact time of assessment, and instructed users to “ignore prompts that could disrupt their work or personal communication” (Morris et al., 2010). Hsieh et al. informed participants that “while they are not required to respond to all the questionnaires, it would be better for the study if they completed as many as possible” (Hsieh, Li, Dey, Forlizzi, & Hudson, 2008).
4.4.4 Questions & Questionnaires

“Care in the design of questions and response formats can greatly reduce the burden by enabling minimal response time and by making the task more pleasant. To date, this important practical issue has not received much attention in the momentary assessment literature.” Stone & Shiffman, 2002

While the literature concerning questionnaire development is significant, the EMA format remains understudied (Stone & Shiffman, 2002). Schwarz writes that responding to any question requires first interpreting the question, then retrieving relevant information, forming a response, mapping this response onto the options provided, and editing the response for reasons of social desirability and self-presentation (Schwarz, 2012). These are steps amenable to design.

What is a good mood, an average relationship or a bad experience? One of the challenges of question design is establishing a shared interpretation of language, intent and experience.21 The ‘pragmatic meaning’ of questions posed in research settings is informed by various factors, including “the purpose of the study,” “the researcher’s affiliation to the content of adjacent questions” and “the nature of the response alternatives” (Schwarz, 2012). Adopting concise and targeted language can therefore support face validity.

Presenting a scale enforces a frame of reference. This can lead to bias, imposing ‘reference periods’ or frequencies (a scale which presents low frequencies implies the reporting of major events for example), invoking our tendency to attempt to complete each category of a scale proportionally across multiple questions and to avoid repeating responses, or producing anchoring and priming effects (Schwarz, 2012).22 The personal use of scales is a particularly difficult bias to counter. This refers to differences of experience and interpretation between respondents; “when Tim answers a 4 …maybe that is the equivalent of a 6 for Jim” (Kahneman & Krueger, 2006). The U-index was introduced as one possible solution to this problem of subjectivity. This momentary measure represents the percentage of time an individual spends in a given (unpleasant) state, in theory reducing the degree of personal sense-making involved in assessment (Kahneman & Krueger, 2006). Others have proposed tailoring “mood queries to an individual’s emotional signature, that is, the range and pattern of each person’s emotions” (Morris et al., 2010).

In the past, researchers often implemented training programs for EMA participants with an eye towards supporting engagement and the collection of valid data. This could include 6 days of “video presentations and real-life situations” (Hormuth, 1986; Perrez et al., 2000). Perrez and Reicherts describe four steps typical of such programs; describing the meaning of underlying constructs, introducing the
use of the ‘pocket computer,’ recording fictional episodes, and performing a trial of the procedure under real life conditions (Perrez & Reicherts, 1996). Conducting user testing and providing instructions to participants may still play a role in achieving the compliance and engagement of users (Christensen et al., 2003; Lanzola et al., 2014).

The scales employed during EMA studies are often designed to capture nominal, ordinal, interval or ratio data as efficiently as possible (Brinkman, 2009). More complex conceptualisation of a domain may require the use of coding schemes. Hormuth describes a scheme for a paper booklet which allows participants to describe their activity in the moment “rather precisely” using four items (location, interactant, activity and conversation) for each of which 15 to 20 response options are provided (Hormuth, 1986). EMA protocols often employ closed-ended scales and questionnaires. This practice facilitates efficient reporting, simplifies analysis and can avoid ethical dilemmas by limiting response options. However, this requires respondents to characterise their thoughts, emotions and behaviours “according to meaningful categories that are often quite abstract,” and can change behaviour by repeatedly presenting lists of categories which influence a subject’s understanding of the domain or serve to remind them of their coping options (Stone et al., 1999).

Mobile devices also permit branching; the dynamic adaptation of scales to previous responses (Christensen et al., 2003; Santangelo et al., 2013). Branching can facilitate refined analysis of inter-social and person-environment interactions (Perrez et al., 2000), and support more engaging patterns of self-report akin to a ‘virtual conversation’ (Lanzola et al., 2014). Branching can also affect data validity. The order of questions can introduce priming effects (Schwarz & Hippler, 1995; Christensen et al., 2003; Kahneman & Krueger, 2006; Schwarz, 2012) and several studies have described attempts by users to take advantage of branching in order to shorten the reporting process (Perrez et al., 2000; Freedman et al., 2006).

Many choices in the design of EMA systems shape both user engagement and data validity. Allowing users to skip questions, for example, might introduce bias and produce less data, while also reducing false reporting rates, revealing flaws in frequently-skipped questions and improving engagement by means of support for users’ autonomy.23

Open-ended questions support greater freedom of expression and allow users to provide additional forms of information and sources of insight, including pertinent features of context (Runyan et al., 2013). Mobile devices enable not just text but multi-media input; photos, audio, video, drawings and more. Although understudied, these means of self-report have the potential to facilitate less structured, more creative, and increasingly conversational modes of interaction which may in turn support
insightful self-disclosure and more meaningful forms of communication (Johnston, 2004; Karapanos et al., 2010; Rains & Keating, 2011).

4.4.5 Reactivity, Habituation & Attrition

Although momentary reporting in daily life is often employed to counter the weaknesses of retrospective assessment, it is not a "direct 'pipeline' into consciousness" but has its own weaknesses pertinent to design (Christensen et al., 2003) (See Section 3.7).

Self-report requires self-reflective awareness; the ability to access relevant information and the willingness to report it (L. F. Barrett & Barrett, 2001). Users can, and are likely to, exercise control over their disclosure. Designers can support the collection of valid data by shaping the context and experience of self-report.24 Researchers conducting a study of smoking behaviours among adolescents (n=562), for example, included non-smokers to avoid the implication that all participants were smoking (Mermelstein et al., 2007). A demo feature allowed users to explain the study to family, friends and teachers, and password-protection was introduced to prevent other students from entering fake data.

Engaging in reflection is not an inherently positive activity (Hormuth, 1986; Perrez & Reicherts, 1987; Siewert, Kubiak, Jonas, & Weber, 2011). That self-focus is a known characteristic of depression, for example, requires designers to consider the kinds of reflection users are asked to engage in (Rude, Gortner, & Pennebaker, 2004); whether 'brooding' or 'reflective pondering' (Moberly & Watkins, 2008), analytical rumination or experiential (mindful) self-awareness (Watkins & Teasdale, 2004), and mal-adaptive or adaptive self-focus (Watkins & Teasdale, 2004). Calvo & Peters suggest designing to support self-compassion; by understanding potential pitfalls, acknowledging the limitations of technology, leaning towards reflective support and allowing for non-absolute categories (Calvo & Peters, 2014).25

The repeated and time-sensitive nature of momentary assessment can lead to reactivity; “the potential for behavior or experience to be affected by the act of assessing it” (Shiffman et al., 2008). Reactivity is associated with increased concept or self-awareness, enhanced “encoding or retrieval of domain-relevant information” and gradual entrainment of participants’ conceptualisation of a domain to match its assessment (Bolger et al., 2003). EMA studies have found both significant (Wilhelm & Perrez, 2013; Robbins & Kubiak, 2014) and insignificant (Hufford, Shields, Shiffman, Paty, & Balabanis, 2002; D. J. Beal & Weiss, 2003) reactivity effects.26 Thirty percent of participants in one EMA study of cocaine-addiction reported that their participation “increased their self-awareness about thoughts, feelings, and behaviors and helped them make decisions that supported abstinence.” Two users also commented that “frequent questioning about cocaine cravings and use could be a trigger to relapse” (Freedman et al., 2007).
It is essential to understand whether we are “tapping a phenomenon as it exists, or as it has been transformed by measurement” (Scollon et al., 2009). Methods for the assessment of reactivity include the use of control groups, querying participants directly, and analysing changes in mean response time, the meaning assigned to scale ratings and the internal consistency of scales over time (L. F. Barrett & Barrett, 2001; D. J. Beal & Weiss, 2003).

The burden of reporting in daily life often leads to high rates of attrition and habituation (Scollon et al., 2009; Santangelo et al., 2013). Beal & Weiss write of their experience of a ‘clear deterioration’ in the quality of momentary mood reporting after the second week of data collection (D. J. Beal & Weiss, 2003), a sentiment mirrored more recently by HCI researchers who have expressed doubt with respect to the possibility of gathering self-reported data in daily life for periods longer than 2 to 4 weeks (Van Berkel, Goncalves, Hosio, & Kostakos, 2017). During another study of time use, participants (n=81) were prompted to provide reports 5 to 7 times a day for 3 non-sequential weeks (Runyan et al., 2013). The average number of responses provided decreased from 18.84 in week 3 to 16.06 by week 8 and to 9.15 by week 14. Eighty percent of users reported greater awareness and 44% expressed “changing how they spent their time.” Reactivity, habituation and attrition pose significant challenges for designers of EMA technologies. These constraints pertain to protocol design, the burden on users, the phenomena of interest, honest disclosure, the nature of reflection, self-awareness and participants’ motivation, sensitivity and desire for change (Perrez et al., 2000; Scollon et al., 2009). Motivated users are more likely to provide data, but also to change their behaviour (Korotitsch & Nelson-Gray, 1999).

4.4.6 Validity in Daily Life

Assessing subjective experience in daily life can support ecological validity but also increases the potential variety of confounding variables. Design for data validity might attempt to account for these effects.

Ancillary data can permit the assessment of validity by facilitating analysis of the time of day or week, features of context such as the weather or location (See Section 2.4.2), extraneous circumstances and semantic descriptions of behaviour, location, social context or psychological antecedents (D. J. Beal & Weiss, 2003). These practices have been described as integrating “satisfaction surveys,” monitoring “participant motivation and quality over time” (Hicks et al., 2010), and building “controls and checks on the subject” into “experience sampling designs” (Hormuth, 1986).

Other variables might be more appropriately assessed prior to or following the reporting period. Pre-sampling questionnaires can capture personality traits, social-desirability bias, technology accep-
tance or experience, motivations, and implicit theories concerning relationships between reported variables or beliefs pertinent to the phenomena of interest (McFarland et al., 1989; Perrez & Reicherts, 1996; D. J. Beal & Weiss, 2003).

Mobile devices provide a variety of unique avenues for the recruitment of participants, including the potential for large-scale public engagement through online app stores (Killingsworth & Gilbert, 2010; Miller, 2012; Trull & Ebner-Priemer, 2013). However, these strategies can also introduce selection biases, exclude those who do not possess a particular variety of mobile device, or attract only motivated and invested participants. Post-sampling assessments can explore usability and user experience concerns, gather participants’ impressions of the validity of their data, and examine other indicators of reliability, validity and reactivity (Perrez & Reicherts, 1996; Fahrenberg et al., 2001). Comparing momentary and retrospective reports might also support a more complete understanding of users’ experience (See Section 3.8).

Several researchers have attempted to combine the advantages of momentary and retrospective reporting protocols (V.-J. Khan, Markopoulos, & IJsselsteijn, 2007; V.-J. Khan et al., 2008; Cherubini & Oliver, 2009). The Day Reconstruction Method (DRM) is the most prominent of these attempts. Completed once at the end of the day, this measure combines reconstruction of the day’s events in diary form with items related to respondents’ emotions, activities and circumstances (Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004). Kahneman et al. however, describe experience sampling as “the gold standard to which DRM results must be compared” (ibid.).

4.4.7 Assessment & Intervention

The combination of assessment and intervention in daily life has been described as one of the most effective means we possess to ‘reduce misery’ and advance wellbeing (Kahneman & Krueger, 2006). Ecological Momentary Intervention (EMI) has been characterised as a “therapist in your pocket” approach with “the potential to revolutionize clinical treatment” (Shiffman et al., 2008). Many features of EMI technologies, including patterns of reflection and feedback, are also relevant to the design of EMA systems (K. E. Heron & Smyth, 2010). In fact, the line between assessment and intervention is often blurred. Momentary assessments can also be employed to tailor interventions (Wichers et al., 2011) and motivate or realise change, by producing ‘external memories’ which guard against stress for example (Perrez & Reicherts, 1987, 1996; Perrez et al., 2000). Not all self-report technologies are driven primarily by the need to gather valid data, but also by the desire to support self-insight, meaning, change, reflection, learning, empowerment, self-determination, and even enjoyment.
4.5 Conclusion | Reflections & Takeaways

The self-report of experience in daily life has often been proclaimed to possess the potential to transform public health, clinical psychology, human-centred research, and design practice. The design of technologies to support these practices is uniquely challenging however, and their adoption has been limited to date. Designers of EMA technologies are tasked with implementing context-appropriate features, navigating divergent modes of use, motivations and outcomes, choosing appropriate questions and sampling protocols, as well as measuring and mitigating bias, reactivity, habituation, attrition and confounding variables, often in pursuit of both the engagement of users and the collection of valid data. This chapter highlights the features of this design space, the dials at the hands of the EMA technology designer which shape the feasibility of these systems’ deployment for public health screening and well-being research. We express, as Miller writes, the need to “take seriously the idea that the people we study are not just passive ‘subjects’ but active ‘participants,’” requiring the acknowledgement of “a genuine continuum between participant and collaborator,” methodology and technology, design and use (Miller, 2012).
Chapter 5

Understanding User Engagement

As we have seen in Chapter 4, designing effective self-report technologies in large part comes down to designing engaging self-report technologies. In fact, the concept of engagement is often doubly relevant to self-report.

The collection of valid data hinges upon the engagement of users, often over significant periods of time (Dunn, Casey, Sheffield, Newcombe, & Chang, 2011). EMA studies typically report a swift decline in reporting practice following “an initial burst of interest” (D. J. Beal & Weiss, 2003; Christensen et al., 2003; Cherubini & Oliver, 2009). In previous design research, for example, the suggestion that pregnant women should “actively track and record their activities” was “met with incredulity, laughter and sometimes derision” (Peyton, Poole, Reddy, Kraschnewski, & Chuang, 2014a).

Secondly, many self-report technologies are designed to inspire engagement with processes and outcomes other than reporting alone (Trull & Ebner-Priemer, 2013). EMA is frequently seen as a means to “actively engage patients in the process of recovery” and increase users’ insight with respect to their own mental health and wellbeing (Wichers et al., 2011). Engagement itself is often considered a characteristic component of wellbeing and the life well-lived (See Section 2.5) (Seligman, 2012).

In fact, there are few designers whose aim is not to engage users. Engagement is a major theme of research across HCI and related fields including artificial intelligence, adaptive computing, learning systems, e-government, social networking, behaviour change and affective computing. Efforts have been made to measure engagement with mobile devices, games, online platforms, conversations, virtual agents, education, healthcare, public discourse and more. However, this is a topic which continues to pose challenges for researchers and designers. What is engagement? What does it mean to be engaged? Which theoretical frameworks might support engagement? Can engagement be measured? Can we design for it? Many have struggled to pin down this nebulous concept;
“It is unclear exactly what to measure and how to measure it and what the important measures are for specific scenarios. It is also unclear how different aspects of user engagement relate to each other.” Attfield, Kazai, Lalmas, & Piwowarski, 2011

“The multidimensional nature of user engagement makes it challenging to measure. While we are very comfortable measuring concrete events, such as the number of errors a user makes when interacting with a system or how long it takes to find the answer to a factual search query, we are less firmly seated when it comes to activities for which there are no visible or physical outcomes.” O’Brien & Toms, 2013

This chapter examines these recurring questions through systematic review.

5.1 Literature Review Methodology

This chapter presents, to the best of the author's knowledge, the first systematic review of engagement across computer science. This review was conducted according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (See Fig 5.1) (Moher, Liberati, Tetzlaff, & Altman, 2009; Kitchenham et al., 2009) and, with the adoption of minimal exclusion criteria, produced a significant and representative corpus with which to consider the question; 'How is the concept of engagement interpreted, understood and measured across the HCI and computer science literature?'

This review corpus was sourced, during the month of May 2014, from the Association for Computing Machinery (ACM) Guide to Computing Literature, the single largest source of computer science literature, containing just short of 2.5 million papers. A search for publications including the term engagement in either title or keyword produced 1,150 results, which were then parsed by a single reviewer to exclude those works outside the remit of the review (See Fig 5.1). Papers were rejected during this initial screening if they;

1. did not focus upon the use of digital technology

and, during a second, full-text assessment for eligibility if they,

1. provided no significant or insightful focus upon, discussion of or application of the concept of engagement

2. took the form of dissertations

3. took the form of books

No additional criteria, such as date or source constraints, were set. Examples of papers excluded during this phase include those focused on automatic transmission engagement in automobiles, civic
engagement with public policy without the mediation of a computer system, or studies of engagement with the STEM (Science, Technology, Engineering and Mathematics) domain. This process produced a final corpus of 351 papers, with respect to which 3 categories of research could immediately be discerned; pertaining to user, student and civil (or public) engagement (See Fig 5.1).

During the review process, a spreadsheet was used to log the relevant features of each paper, ensuring consistency and replicability. This log structure incorporates coding dimensions derived from the aims of the review, initial assessment of the engagement literature, and the research methods literature (C. Hart, 2001; Grix, 2010; Fink, 2013) as well as the PRISMA guidelines (Moher et al., 2009).

- Administrative Coding (title, author(s), date published, publication type and title, key words)
• Engagement Coding (definition, theories, attribute or process construal, number of states, engagement focus, social emphasis, passive or active, negative perspectives, application context, users targeted, platform, measures applied, engagement synonyms)

• Methods and Study Quality Coding (sample size, sample demographics, experimental design, sampling procedure, statistical techniques, nature of the reported data, study aims, originality, study description, reviewer notes)

This coding scheme informed a thematic analysis of the final corpus, leading to a detailed mapping of the landscape of engagement research. Analysis of this corpus provides immediate insight. As expected, interpretation of engagement is highly diverse. Conferences papers (213), journal articles (127), book chapters, posters and articles (11) all feature discussion of the concept. The publication of engagement research is increasing exponentially, at a rate of approximately 28% per year. While this trend likely coincides with an increase in user experience research, it illustrates the growing body of scholarship on this topic.

5.2 What Is Engagement?

The need to understand users and their interactions with technology first motivated a turn towards the study of user experience within HCI (McCarthy & Wright, 2004). This has been followed by a turn towards user engagement (Bardzell et al., 2008). McCarthy & Wright comment that our efforts to understand user experience are complicated by our own immersion in experience and a conventional assumption of consistent interpretation (McCarthy & Wright, 2004). Engagement is likewise a feature of our day-to-day language, and often a concept whose potential for entanglement only emerges once attempts are made at its definition and measurement. A measure of this complexity can be discerned by examining definitions of the term.
5.2.1 Definitions of Engagement

The 351 papers of this corpus featured 102 definitions of engagement. Several conceptions were referenced by multiple publications, the single most-cited being that proposed by Sidner et al., encountered on 6 occasions;

“By engagement, we mean the process by which two (or more) participants establish, maintain and end their perceived connection. This process includes: initial contact, negotiating a collaboration, checking that other is still taking part in the interaction, evaluating whether to stay involved, and deciding when to end the connection.” C. L. Sidner, Kidd, Lee, & Lesh, 2004 (emphasis added)

This definition positions engagement as a component of social connection. Any explicit definition constitutes certain assumptions on the part of the author. Sidner et al., in this instance, explicitly cast engagement as a process composed of three distinct phases, a beginning, a period of sustentation and an end. These phases are then further divided into a series of user actions. This framing implicitly places the definition within the context of a conversation between at least two agents, where the user is cast as an active and receptive participant and engagement as a continuous, synchronous process with a clearly defined beginning and end. Established from the user’s perspective and distinguishing discrete components of interaction, this definition lends itself to the development of autonomous and task-focused systems, rather than to a single user’s engagement with a screen-based interface for example.

Quesenbury links engagement to the qualities of an interface;

“the degree to which the tone and style of the interface makes the product pleasant or satisfying to use.” Quesenbury, 2002 (emphasis added)

O’Brien & Toms describe engagement as a quality of user experience with multiple component features;

“Engagement is a quality of user experiences with technology that is characterized by challenge, aesthetic and sensory appeal, feedback, novelty, interactivity, perceived control and time, awareness, motivation, interest, and affect.” O’Brien & Toms, 2008 (emphasis added)

Zyngier divides student engagement into behavioural, emotional and cognitive components;

“Student engagement is a mix of several components 1) Behavioural: persistence and participation, 2) Emotional: interest, value and valence, and 3) Cognitive: motivation, effort and strategy.” Zyngier, 2008 (emphasis added)
Douglas & Hargadon supply a definition of engagement with a distinct hedonic inclination, tied to immersion, narrative flow and schema theory;

“The pleasures of immersion stem from our being completely absorbed within the ebb and flow of a familiar narrative schema. The pleasures of engagement tend to come from our ability to recognize a work's overturning or conjoining conflicting schemas from a perspective outside the text, our perspective removed from any single schema.” Douglas & Hargadon, 2000 (emphasis added)

Dobrian et al. describe engagement as a proxy for involvement and interaction;

“Qualitatively, engagement is a reflection of user involvement and interaction.” Dobrian et al., 2013 (emphasis added)

Jaimes et al. define engagement in the context of social media, employing the language of captivation and motivation;

“Engagement defines the phenomena of being captivated and motivated: engagement can be measured in terms of a single interactive session or of a more long-term relationship with the social platform across multiple interactions. Thus, social media engagement is not just about how a single interaction unfolds, but about how and why people develop a relationship with a platform or service and integrate it into their lives.” Jaimes, Lalmas, & Volkovich, 2011a (emphasis added)

Laurel ties engagement directly to an emotional state;

“Engagement refers to the (emotional) state of mind the user must attain to enjoy the representation, that is, a willing suspension of disbelief.” Laurel, 1991 (emphasis added)

This synopsis highlights the diverse language brought to bear on the topic of engagement and the importance of context in weighing the value of any one definition. Engagement has been studied across many contexts, and with respect to numerous forms of technology (See Table 5.1).

However, 65% of publications which address engagement do not provide a definition. This likely contributes to the persistent ambiguity surrounding the term. To more fully grasp this diversity of interpretation we turn to the theoretical foundations of engagement.
5.3 Engagement Theory

In striving to understand engagement, researchers have looked to, and developed, a broad range of theory. This sample of the literature contained 372 theoretical frameworks (See Table 5.2), nets cast to ‘rationalise’ and ‘explain’ human-computer interaction (Popper, 1959). Certain frameworks focus primarily on engagement (O’Brien & Toms, 2010a). Others on related concepts such as immersion (Carrigy, Naliuka, Paterson, & Haahr, 2010) and involvement (Webster & Ahuja, 2006).

5.3.1 Levels of Theory

Discerning categories of engagement theory supports the distinction of approaches to and motivations for interpretation of the concept. Theories are often described as focused on either micro (actor-focused) or macro (structure-focused) processes. Flow theory (Cowley, Charles, Black, & Hickey, 2008), cognitive load theory (Oviatt, Swindells, & Arthur, 2008) and social presence theory (J. J. Choi, Kim, & Kwak, 2014), for example, tend to construe engagement in terms of the nature and distribution of an individual’s conscious focus. This facilitates the analysis of individual episodes of interaction and the inference of engagement as a state of experience. Fourteen percent of the engagement literature references flow theory (See Table 5.2) (Webster & Ahuja, 2006; Seah & Cairns, 2008). This theory posits a state of optimal and enjoyable experience characterised by a tractable challenge, immersion, control, freedom, clarity, immediate feedback, temporal insensitivity and change to one’s sense of identity (Cowley et al., 2008). This framework provides a means of measuring flow, often employed as a synonym for engagement, through its component parts, as well as a target for would-be interaction designers; the
flow state. This factored model presents engagement (flow) as demarcated and attainable through a balance of factors including challenge and skill.

In contrast, theories of motivation (Seddon, Skinner, & Postlethwaite, 2008), self-determination theory (Wiebe, Lamb, Hardy, & Sharek, 2014) and narrative theories (Douglas & Hargadon, 2000) tend to support higher-level analysis, reflecting aggregate, socio-structural and temporal factors. Schema theory has been employed in the analysis of flow, engagement and immersion in traditional and interactive narratives for example (Douglas & Hargadon, 2000). Macro perspectives also feature in reflection on the use of mobile media technologies (Lai, 2014), ethnographic analysis of users’ *environmental engagement* through in-car GPS (Global Positioning System) systems (Leshed, Velden, Rieger, Kot, &

<table>
<thead>
<tr>
<th>Theory</th>
<th>Mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow Theory</td>
<td>50</td>
</tr>
<tr>
<td>Motivation Theory</td>
<td>16</td>
</tr>
<tr>
<td>O’Brien and Tom’s Model of Engagement</td>
<td>13</td>
</tr>
<tr>
<td>Sidner et al’s Model of Engagement</td>
<td>9</td>
</tr>
<tr>
<td>Presence Theory</td>
<td>9</td>
</tr>
<tr>
<td>Play Theory</td>
<td>7</td>
</tr>
<tr>
<td>Self Determination Theory</td>
<td>6</td>
</tr>
<tr>
<td>Independent Television Commission Sense of Presence Inventory</td>
<td>6</td>
</tr>
<tr>
<td>ITICSE Engagement Taxonomy</td>
<td>6</td>
</tr>
<tr>
<td>Technology Acceptance Model</td>
<td>5</td>
</tr>
<tr>
<td>Social Presence Theory</td>
<td>5</td>
</tr>
<tr>
<td>Immersion Theory</td>
<td>5</td>
</tr>
<tr>
<td>Activity Theory</td>
<td>4</td>
</tr>
<tr>
<td>Involvement Theory</td>
<td>4</td>
</tr>
<tr>
<td>Cognitive Load Theory</td>
<td>4</td>
</tr>
<tr>
<td>User Engagement Scale</td>
<td>3</td>
</tr>
<tr>
<td>Brown and Cairn’s Model of Engagement</td>
<td>3</td>
</tr>
<tr>
<td>Media Equation Theory</td>
<td>3</td>
</tr>
<tr>
<td>Peter et al’s Model of Engagement</td>
<td>3</td>
</tr>
<tr>
<td>Game Engagement Questionnaire</td>
<td>3</td>
</tr>
<tr>
<td>Active Learning Theory</td>
<td>3</td>
</tr>
<tr>
<td>Social Capital Theory</td>
<td>3</td>
</tr>
<tr>
<td>Richness Theory</td>
<td>2</td>
</tr>
<tr>
<td>Engagement Sampling Questionnaire</td>
<td>2</td>
</tr>
<tr>
<td>Immersive Tendencies Questionnaire</td>
<td>2</td>
</tr>
<tr>
<td>Michalowski et al’s Model of Engagement</td>
<td>2</td>
</tr>
<tr>
<td>Kendon’s F-Formation Theory</td>
<td>2</td>
</tr>
<tr>
<td>Ermi and Mayra’s Model of Immersion</td>
<td>2</td>
</tr>
<tr>
<td>etc.</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.2: Engagement & Theory
Sengers, 2008), and in Borgmann’s critique of the pattern of modern life as “one of disengagement, in
which technological devices might relieve us of tedious effort but do so at the expense of our engage-
ment with the world” (Leshed et al., 2008; Dindler, Iversen, & Krogh, 2011).

5.3.2 Cognitive, Emotional and Behavioural Engagement

Is engagement cognitive, emotional or behavioural? Several authors make such distinctions (Zyngier,

Interpretations of engagement as **cognitive** in nature often focus on conscious activities or concepts
such as effort (Ilsas Sedeno et al., 2013; J. C.-Y. Sun, 2014), energy (Birnholtz, Hancock, & Retelny,
2013), awareness and attention (Ilsas Sedeno et al., 2013). Such interpretation renders engagement
more easily quantified and is often accompanied by a focus on process and strategy (Bouta & Retalis,
2013; Islas Sedano et al., 2013), as in the case of Fredricks et al’s **cognitive engagement scale**, which
denotes “an individual’s voluntary efforts to understand and master challenging tasks,” for example (See
J. C.-Y. Sun, 2014).

Engagement has also been described as “primarily understood as an emotion” (Laurel, 1991). Def-
initions of **emotional** or affective engagement tend to emphasise the subjective nature of experience,
encompassing identification, belonging, values, attitudes and emotions (Ilsas Sedeno et al., 2013), re-
flected in expressions of interest, boredom, achievement orientation, values and feelings (Bouta & Re-
talis, 2013). Interpreting engagement in these terms allows researchers to bring theories of affect and
emotion to bear on engagement. One such interpretation describes engagement with respect to the
circumplex model of emotion — as a state of high valence and arousal (Latulipe, Carroll, & Lottridge,
2011). Another locates engagement between boredom and stress “within a framework of arousal states”
(Churchill, 2010).

A **behavioural** construal of engagement emphasises action and participation, promising more ob-
jective measurement (Asteriadis, Karpouzis, & Kollias, 2009; Bouta & Retalis, 2013; Islas Sedeno et al.,
2013). Bianchi-Berthouze et al. express the belief that the cognitive focus of many theories of engage-
ment neglects an essential physical component (Bianchi-Berthouze, Kim, & Patel, 2007). Studies of
gameplay using the GuitarHero controller reveal a possible correlation between player engagement and
physical movement, which appears “not only to increase the players’ level of engagement but also to
modify the way they get engaged” (Bianchi-Berthouze et al., 2007).

Conceptions of engagement can be divided along many lines, “academic, cognitive, intellectual, in-
stitutional, emotional, behavioural, social, and psychological” for example (Ting, Cheah, & Ho, 2013).
These distinctions are not based on objective knowledge of engagement as a universal neurophysiological or social phenomenon but are made to facilitate a particular form of context-dependent analysis or design. These motivations reflect the ontological perspectives brought to bear on engagement (See Section 5.5).

5.3.3 Trait, State or Process Characterisations

One useful means of distinguishing between conceptions of engagement is to examine trait, state and process-based interpretations. These characterisations are almost always made implicitly, and so merit closer examination.

Engagement can be interpreted as **trait** based; a stable characteristic of persons or systems, a propensity to engage or be engaged. *Cognitive absorption*, for example, has been described as a user’s “propensity to become absorbed in the activities around using a computer,” in contrast to *immersion*, construed in this instance as a comparatively dynamic state (Seah & Cairns, 2008). *Within-person process* and *individual-difference structural* theories of cognition have long been the subject of debate in personality psychology (Fleeson, 2007). However, the role of personality with respect to engagement featured only rarely within this corpus (Cowley et al., 2008; Goldberg, Sottilare, Brawner, & Holden, 2011; Bixler & D’Mello, 2013). Improved knowledge of individual patterns of engagement might support the design of personalised, and more engaging, experiences (Barco, Albo-Canals, & Garriga, 2014).

Engagement is most frequently characterised as a variable **state**, ascribed to the user (O’Brien & Toms, 2008), a **system** (Bohus & Horvitz, 2009c) or to **interaction** itself (Deray & Simoff, 2012). These conceptions tend to emphasise perception, reflection and a responsive, dyadic form of interaction. A key feature of any discrete variable is the number of states it can adopt. The design of Pêle-Mêle, a multi-party video communication system, explored this choice with respect to engagement, employing measures of users’ presence, motion and distance to determine the system’s state of engagement during use (Gueddana & Roussel, 2006). The view that technologies should be designed to support variable degrees of engagement, and smooth transitions between each, merits wider adoption. Casting engagement as a dynamic state, whether discrete or continuous, is a reductive stance often adopted to facilitate measurement and modelling. This perspective frames engagement as a finite resource, a form of conscious currency to be managed by human and machine in interaction.

Although engagement is most often cast as a state, there is a discernible consensus across the literature concerning its functioning as a **process** (O’Brien & Toms, 2008). The processual nature of engagement is often lost however in its operationalisation as a more easily measured discrete state. Construing
engagement as a process facilitates the analysis of change over time and between periods of interaction. One hybrid view describes a \textit{process of transition} between different \textit{modes} of engagement, from physical to contemplative (Dindler et al., 2011).

5.3.4 Prefixation

Engagement is a highly versatile concept whose definition often entails realising a reduction in scope. A number of authors have chosen to do so by applying an appropriate prefix. Chilukuri & Indurkhya, for example, describe \textit{narrative engagement} as a concept comprising narrative understanding, attentional focus, emotional engagement and narrative presence (Chilukuri & Indurkhya, 2011). Churchill writes “how about ‘vongagement’ — for voluntary engagement? Or ‘invitagement’ for invited engagement? Or ‘partagement’ for participatory engagement?” (Churchill, 2010). Such distinctions are often key to the value of engagement as a concept.

5.4 The Ecology of Engagement

![An HCI Ecology](image)

Figure 5.3: An HCI Ecology

Advancing engagement research requires an understanding of conceptual distinctions. In addition to examining the theoretical frameworks employed by researchers and designers, we propose turning to the ontological framing of engagement, expressions of which are rarely explicit. We locate conceptions
of engagement within the framework of an HCI ecology composed of state, agent and interaction (See Fig 5.3).

5.4.1 The Engaged State

Engagement is often positioned at the level of individual conscious experience, whether described as a state (Laurel, 1991; M. Chen, Kolko, Cuddihy, & Medina, 2011; Goldberg et al., 2011; Sundar, Oh, Bellur, Jia, & Kim, 2012; Houtkamp, Toet, & Bos, 2012) or perceived quality of experience (Baeza-Yates & Lalmas, 2012; Lehmann, Lalmas, Yom-Tov, & Dupret, 2012; McCay-Peet, Lalmas, & Navalpakam, 2012; O’Brien & Toms, 2013). These characterisations are subject to degrees of abstraction. State-based interpretations of engagement range from highly-reductive transactional conceptions (Gueddana & Roussel, 2006) to complex representations of subjective experience (Sundar et al., 2012). Affective, cognitive and conative (behavioural) states of engagement often reflect an inherently rationalist epistemology (Zyngier, 2008; Abadi, Staiano, Cappelletti, Zancanaro, & Sebe, 2013; Islas Sedano et al., 2013).

Engagement is also, and often simultaneously, described as a qualitative character of experience (Marsh et al., 2008; Corrigan, Peters, & Castellano, 2013; Y. H. Kim, Kim, & Wachter, 2013). This is a phenomenological framing, entailing interpretation of engagement as a distinct shade of conscious experience. Such instances of subjective sensation have been described as ‘qualia’ (latin: meaning ‘of what kind’) (Dennett, 1988). These states are ‘intentional,’ in the sense that they are ‘about something’ (Dennett, 1971). Definitions of engagement often embody an ‘object,’ a focus, an intentionality.

5.4.2 The Engaged Agent

Engagement is also frequently accorded meaning at the level of users and systems (Hwang & Thorn, 1999; Webster & Ahuja, 2006; Dow, 2007; Poggi, 2007; Castellano, Pereira, Leite, Paiva, & McOwan, 2009; Schoenau-Fog, 2011). When we speak of engagement in these terms, we refer not to action alone or to a state in isolation but often invoke valence, desire, commitment, participation, volition, attention, immersion and motivation.

Understanding the Engaged User

Efforts to define the engaged user may consider predispositions, expectations, motivations, mood, age, gender and so on. Learning systems research often features attempts to model systems and users (Naps et al., 2002; Tootell, Plumb, Hadfield, & Dawson, 2013; Kucirkova, Messer, Sheehy, & Fernández Panadero, 2014). One student may disengage because she “finds the work too easy,” another “may be
capable of doing the work but lacks confidence and feels too anxious about failure,” and another “may not have the required skills but is wary of” using a system designed to help “because she has learned not to expect useful assistance from peers, parents or even teachers” (C. R. Beal, Qu, & Lee, 2006).

Motivation

Characterisations of engaged agents often incorporate related concepts such as motivation — additional abstractions of self and experience (Tashiro & Dunlap, 2007; Cocea & Weibelzahl, 2007; O’Brien & Toms, 2010a; Liu, Alexandrova, & Nakajima, 2011; Y. H. Kim et al., 2013). Motivation has been defined as a dynamic process in which sustained effort is applied to pursue goals which satisfy needs “subject to cognitive processes and set against values” (Seddon et al., 2008). Motivational factors are believed to include reasons for action (needs and values), interest, a conducive social context, effective feedback and a sense of agency arising from choice, control and an expectancy of success (self-efficacy and self-worth) (Seddon et al., 2008). The process of engagement has also been described as a ‘feedback loop’ in which experience with a task shapes “the more state-like elements of self-efficacy and motivation which in turn influences the user’s desire to re-engage” (Wiebe et al., 2014). The Mobile User Engagement Model (MoEn) associates engagement with functional (efficiency, ease of use, saving time), hedonic (fun, enjoyment, pleasure) and social (desire to connect and share with others) motivations (Y. H. Kim et al., 2013).

5.4.3 Engaged Interaction

Engagement has also been described as a feature of interaction itself rather than a character of states or agents (Davies, 2002; C. L. Sidner, Lee, Kidd, Lesh, & Rich, 2005; Deray & Simoff, 2012; Dobrian et al., 2013). Interaction can refer to actions and relationships between states, agents and objects, in the moment or over time. Within the engagement literature, interaction is often employed as a direct proxy for engagement, marrying engagement to a scale of use, the unit of which is interaction. This is a measure of quantity not quality. Other interpretations associate degrees of engagement with different forms of interaction. Mallon & Webb’s framework for evaluating engagement in narrative multime dia describes interaction as a dimension of engagement which incorporates skill-based interaction, the relevance of dialogue and causality, the illusion of intelligence and invisibility of the medium (See Schuurink, Houtkamp, & Toet, 2008). Interaction has also been described as “an essentially reactive activity” in contrast to engagement which “has an affective component” (Turner, 2010). While engagement has been viewed as a product of interaction (O’Brien & Toms, 2008; Sundar et al., 2012; Deray & Simoff,
2012), it has also been thought so broad a notion that it is “most productive to consider engagement as a perspective on interaction, rather than a clearly defined entity” (Dalsgaard, Dindler, & Halskov, 2011).

An HCI ecology provides a scaffolding for conceptions of engaged states, experiences, agents and interactions, whose distinction is often implicit within the literature. Attempts to measure engagement further reflect these themes.

5.5 Measuring Engagement

The conception and measurement of engagement are interlinked and contingent upon the motives of designers and researchers. From this corpus we discern three distinct motivations for the adoption of engagement:

1. **Basic Research** Understanding Engagement
2. **Design Practice** Designing Engagement
3. **Systems Development** Implementing Engagement

This first category comprises research focused on the development of generalised theories and models of engagement, often embodying psychological and social factors. This work goes beyond engagement with a particular designed object or process, studying its universal potential. Examples include extensive models of user engagement (O’Brien & Toms, 2008), socio-technological critique (McCarthy & Wright, 2004; Leshed et al., 2008) and sociological analysis (Goffman, 1978).

The adoption of engagement has also been motivated by design practice. Engagement has been employed as a lens through which to analyse users’ experiences with technology probes, and as a concept in user experience design. Patterns of engagement have been used to provide actionable insight into the design of systems including mobile games (Fischer & Benford, 2009), learning technologies (LeeTiernan & Grudin, 2003) and websites (Colbert & Boodoo, 2011).

A third motivation rests on the capacity of engagement to mediate human behaviour and communication. Developers of autonomous systems have leveraged the role of engagement in interaction, approaching the concept with an eye towards synthesis, not just analysis. This approach entails operationalising (translation from conception to signal), modelling (representation within a logical framework), and connecting measures of engagement to appropriate autonomous actions, as in research concerning cognitive assistance during driving (Cai & Lin, 2012), the behaviour of a robot bartender (Foster, Gaschler, & Giuliani, 2013), keystroke analysis during essay writing (Bixler & D’Mello, 2013), atten-
tion to information on-screen (Asteriadis et al., 2009) and conversational agents (Nakano & Yamaoka, 2009; Xu, Li, & Wang, 2013).

5.5.1 The Practice of Measurement

<table>
<thead>
<tr>
<th>Measure</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaire</td>
<td>124</td>
</tr>
<tr>
<td>Behaviour Logging</td>
<td>69</td>
</tr>
<tr>
<td>Observation</td>
<td>44</td>
</tr>
<tr>
<td>Task Outcomes</td>
<td>26</td>
</tr>
<tr>
<td>Interview</td>
<td>23</td>
</tr>
<tr>
<td>Eye Tracking</td>
<td>19</td>
</tr>
<tr>
<td>Discussion</td>
<td>10</td>
</tr>
<tr>
<td>Electroencephalogram (EEG)</td>
<td>10</td>
</tr>
<tr>
<td>Audio Signal Analysis</td>
<td>8</td>
</tr>
<tr>
<td>Galvanic Skin Response (GSR)</td>
<td>8</td>
</tr>
<tr>
<td>Facial Analysis</td>
<td>7</td>
</tr>
<tr>
<td>Review</td>
<td>6</td>
</tr>
<tr>
<td>Head Pose</td>
<td>5</td>
</tr>
<tr>
<td>Physical Motion Tracking</td>
<td>5</td>
</tr>
<tr>
<td>Body Pose</td>
<td>4</td>
</tr>
<tr>
<td>Presence Detection</td>
<td>3</td>
</tr>
<tr>
<td>Electrocardiogram (ECG)</td>
<td>2</td>
</tr>
<tr>
<td>Heart Rate</td>
<td>2</td>
</tr>
<tr>
<td>Facial Electromyography (fEMG)</td>
<td>1</td>
</tr>
<tr>
<td>Gesture Tracking</td>
<td>1</td>
</tr>
<tr>
<td>Heat Flux (HF)</td>
<td>1</td>
</tr>
<tr>
<td>Pen Pressure</td>
<td>1</td>
</tr>
<tr>
<td>Pupil Diameter</td>
<td>1</td>
</tr>
<tr>
<td>Systolic Blood Pressure (SBP)</td>
<td>1</td>
</tr>
<tr>
<td>Supraorbital Thermal Imaging</td>
<td>1</td>
</tr>
<tr>
<td>Temperature Monitoring</td>
<td>1</td>
</tr>
<tr>
<td>No Engagement Measure</td>
<td>106</td>
</tr>
</tbody>
</table>

Table 5.3: Engagement & Measurement

The variety of methods applied to the measurement of engagement speaks to the versatility of the concept (See Table 5.3). Subjectivity-oriented measures tend to reflect the perceptual and experiential nature of many conceptions. Objectivity-oriented measures embody more reductive interpretations, often in the search for actionable data. The literature features many examples of both approaches to measurement including questionnaires, interviews, annotation, measures of outcomes and digital behaviour logging.
Subjectivity-Oriented Approaches

Subjectivity-oriented measures include observation, questionnaires, interviews and other forms of self-report. As for any methodology, the choice of these methods entails compromise. These approaches support rich description coloured by subjectivity, cognition, emotion and memory.

‘Subjective’ Methods

Questionnaires are the most frequently employed measure of engagement (See Table 5.3). They are easily applied and can support detailed analysis (Webster & Ahuja, 2006; Rozendaal, Keyson, & Ridder, 2008; Brockmyer et al., 2009; O’Brien & Toms, 2010a; Seok & DaCosta, 2012). However, they are also often retrospective in nature and yield a personally-motivated and reconstructed perspective (Latulipe et al., 2011).

The development of the User Engagement Scale (UES) is one of the most thorough attempts to devise an understanding of user engagement to be found within the literature (O’Brien & Toms, 2010a). This questionnaire was informed by an initial conceptual framework for user engagement produced following analysis of interviews conducted with 17 participants concerning their experiences with online shopping, searching, video games and distance learning (O’Brien & Toms, 2008). This framework depicts engagement as a process comprised of four stages; a point of engagement, a period of sustained engagement, disengagement and re-engagement, with respect to a system that is “effective, efficient and satisfying.” The UES was further informed by literature review, leading to the description of engagement as;

“both a process and product (Kappelman, 1995) of interaction; its intensity may change over the course of an interaction (Said, 2004) depending on the combination of users’ needs, goals, emotions, actions, and thoughts, or the format (Chapman, 1997; Jacques, 1996), visual presentation, and organization of the computer interface (Quenbury, 2003).”
O’Brien & Toms, 2010a

An initial evaluation study, comprising 440 participants’ responses concerning an online shopping experience, reduced the UES to 33 items distributed across 6 principal factors; Focused Attention, Perceived Usability, Aesthetics, Endurability, Novelty and Felt Involvement (O’Brien & Toms, 2010a). A second study, during which 802 participants completed the revised scale with respect to an online book purchase, identified relationships between these factors. Aesthetics was found to predict perceived usability, focused attention predicted felt involvement, and felt involvement predicted endurability. Sub-
sequent analyses led to further refinements of this model (O’Brien & Toms, 2010b, 2013; Wiebe et al., 2014).

Experience Sampling Methods (ESM) have the potential to capture changes in users’ engagement over time through the repeated self-report of experience in the moment. The Engagement Sampling Questionnaire (ESQ) consists of demographic, pre-experience and in-experience components, focused on user objectives, activities, accomplishment and affect (Schoenau-Fog, 2011). Questions posed in-experience include a Likert-scale based measure of continuation desire and a number of open-ended questions. The ESQ has been applied as part of an iterative game design process (Schoenau-Fog, Birke, & Reng, 2012) and in the development of the experiential game Aporia (Bevensee, Dahlsgaard Boisen, Olsen, Schoenau-Fog, & Bruni, 2012). An ESM approach was also employed to understand users’ engagement with an SMS-based pervasive mobile game (Fischer & Benford, 2009). These measures construe engagement as a momentary concept, a feature of immediate experience, in contrast to retrospective questionnaires for example, which more often reflect multidimensional interpretations.

A wide variety of methods have been employed to capture and describe the subjective experience of engagement. An ethnographic approach informed the development of a 5-phase model of user engagement with public displays; comprising receptiveness, interest, evaluation, engagement and disengagement (Glasnapp & Brdiczka, 2009). Qualitative analysis identified immediate conversation, creative production and reflective observation as dimensions of engagement with research videos (Caglio & Buur, 2012). And, persona development has also been used to explore children’s social interaction and engagement with characters in a virtual school environment (Hall, Jones, & Hall, 2008).

Inferring Engagement from ‘Subjective’ Data

An important distinction among subjective measures is the degree to which a-priori framings bias the inference of engagement. Questionnaires, for example, are often deductive rather than inductive in nature and invite closed-ended responses. These choices are highlighted in the creation of engagement scales (Webster & Ahuja, 2006; Rozendaal et al., 2008; Brockmyer et al., 2009; O’Brien & Toms, 2010a; Seok & DaCosta, 2012). The Game Engagement Questionnaire (GEQ), for example, interprets engagement as a “generic indicator of game involvement” incorporating immersion, presence, flow, psychological absorption and dissociation (Brockmyer et al., 2009). While the User Engagement Scale (UES) measures focused attention, perceived usability, aesthetics, endurability, novelty and felt involvement (O’Brien & Toms, 2010a).

Measures of engagement are context-dependent. O’Brien & Toms, for example, state that the UES...
“incorporated effectiveness and satisfaction, but not efficiency. This makes sense since an engaged user may not be concerned about how much time they are spending” but it means that “some usability metrics, including time spent performing a task or examining a screen, may not be applicable” (O’Brien & Toms, 2010a). Face validity can often indicate whether the questions posed by a particular scale relate coherently to the context of measurement.5

Open-ended methods can support richer analysis although lack the reliability and ease of application of questionnaires. Interviews can reveal ‘indications of engagement’ for example, “reported as ‘awareness’, ‘involvement’, ‘immersion’ and so forth.” Instead of asking “were you engaged?,” participants can be asked to describe their experience (Swift, Gardner, & Riddell, 2010).

5.5.3 Objectivity-Oriented Approaches

Objectivity-oriented measures attempt to infer engagement without recourse to direct questioning or human involvement. This includes such techniques as logging behaviour and interaction, psychophysiological measures and audio and visual analysis. The advantage of these approaches lies in their ease of application, limited disruption of experience and minimal imposition of user burden. We describe these measures as oriented towards objectivity, given their approximation of engagement through its observable features. However, all such measures are themselves imbued with a certain subjectivity, introduced through users’ awareness of observation and researchers’ experimental and interpretive choices.

`Objective’ Methods

The choice of engagement measure has been described as a trade-off between **scale of data capture** and **depth of understanding** (Baeza-Yates & Lalmas, 2012). A distinguishing feature of these measures is therefore the complexity of the data from which engagement is inferred.

The simplest of measures adopt single variables as proxies for engagement, such as the number of mouse clicks on a web-page. These measures are often employed when access to users is limited, as in the case of e-publications (R. S. Lee, Liu, Yeung, Sin, & Shum, 2009), email spam exposure (Dasgupta, Punera, Rao, & Wang, 2012), blogs (Hennig et al., 2013), online networks (Lehmann, Lalmas, Baeza-Yates, & Yom-Tov, 2013), online multitasking (Lehmann, Lalmas, Dupret, & Baeza-Yates, 2013), search engines (Song, Shi, & Fu, 2013) and analysis of the effects of network links (Yom-Tov et al., 2012), video quality (Dobrian et al., 2013) and absence time (Dupret & Lalmas, 2013) on engagement.

Behaviour logging approaches support the assessment of engagement at significant scales. Lehmann et al. collected data from two million users across 80 websites, according to three categories of engage-
ment metrics; popularity, activity and loyalty (Lehmann et al., 2012). This data allowed the authors to define five user types based on the number of days which included a visit to a site, to calculate the proportion of each user group per site and to cluster sites with similar user groups, extricating patterns of engagement (Lehmann et al., 2012).

Another broad category of measures which have been used to infer engagement are psychophysiological. Electrocardiography (ECG), heat flux (HF) and electroencephalography (EEG) measures have all been employed as proxies for engagement (Belle, Hobson, & Najarian, 2011). Features of the signals produced by these methods enabled a binary classifier of engagement to be trained on data captured while eight subjects viewed two 20-minute sets of video clips, one deemed engaging, the other the opposite (Belle et al., 2011). A model based on the EEG signal alone achieved rates of 86%+ for accuracy, precision and recall metrics. ECG and HF models combined achieved only slightly lower rates and all three combined achieved greater than 90% accuracy, precision and recall. EEG measures have also been used to infer music engagement without interrupting the listening experience (n=13) (Blankertz et al., 2010). Predictive models, based on EEG and motion-tracking data, achieved 67% engagement classification accuracy across-subjects and 85% accuracy within-subjects.

Attempts have also been made to infer engagement from audio and video data. Ishii & Nakano measured conversational engagement between a user and a Wizard-of-Oz controlled virtual agent in the role of a mobile phone salesperson by examining eye-tracking data, speech transcriptions, the virtual agent’s gaze and gestures as well as observed and user-reported judgements of user-attitude (Ishii & Nakano, 2008). Analysis of this data revealed that engaged users behaved more cooperatively, redirecting their attention to a focused object when referred to by the agent (Ishii & Nakano, 2010). Further research, incorporating head pose and topic change information, led to the development of a binary classification of engagement with an accuracy rate of 77.8% (using 10-fold cross validation) and a three-class model with an accuracy rate of 88.75% (Ooko, Ishii, & Nakano, 2011). Proponents of these approaches have reported promising results. However, many studies have taken place in laboratory environments, narrow contexts of use and with small numbers of users (Sanghvi et al., 2011; Cai & Lin, 2012; Bixler & D’Mello, 2013; Foster et al., 2013; Xu et al., 2013), and so it is difficult to generalise initial findings.

Inferring Engagement from ‘Objective’ Data

Data from ‘objective’ sources offers several advantages. It can be captured at large scales without human involvement, has the potential to be gathered in an ambulatory fashion and supports the action of autonomous systems. Reductive data can also lead to shallow interpretation however, false attribu-
tions of causality, misleading time-scale responsivity and invalid generalisation. If a user spends a long

time on a web-page, are they engaged or simply confused? (Birkett, Galpin, Cassidy, Marrow, & Nor-
gate, 2011). Time spent searching for information can be “involving, but not in a pleasant way …just
inefficient and frustrating” (Colbert & Boodoo, 2011).

Measures are not always as ‘objective’ as we may assume. The ground truth against which these
methods are evaluated is often derived by self-report, annotation or expert ratings (Sanghvi et al., 2011;
Bonin, Bock, & Campbell, 2012; Xu et al., 2013). Annotated gameplay video was employed as the
ground truth for computer-vision based models of affective posture and body motion in a study of en-
gagement during chess games between children and the iCat robot (Sanghvi et al., 2011). When asked,
annotators revealed, perhaps counter-intuitively, that they had associated less movement and an upright
or backward leaning posture with higher levels of engagement, and patterns of continuous movement
and a forward leaning posture with lower levels of engagement. Another group of researchers, organis-
ing the annotation of individual and group engagement within a video corpus, imposed no constraints
on observers, no time scales for annotation and no definition of engagement (Bonin et al., 2012). When
annotators were asked to describe their process, they stated that they concentrated mainly on “whether
the subjects were speaking or not and on the body pose” to determine an individual’s level of engage-
ment (Bonin et al., 2012).

The perceptual and experiential nature of engagement has significant implications for its measure-
ment, regardless of the methodology employed. We possess agency with respect to affective display, or
perhaps more appropriately in this case engagement display. Although “an entity may have an interest
in something, this is to be differentiated from the action of showing interest in it” (Peters, Castellano, &
de Freitas, 2009). This is particularly important when it is likely for participants to control their display
of engagement, due to social facilitation, inhibition or loafing effects for example (Liu et al., 2011).

While the act of observation alone can influence the display of engagement, many means of eval-
uation also interrupt experience more overtly (LeeTiernan & Grudin, 2003). A distinction between im-
plicit, “non-intrusive, pervasive and embedded” and explicit measures is valuable in this respect (Cor-
rigan et al., 2014). One study of human-robot interaction, for example, supplied users with a ques-
tionnaire every minute in addition to measuring the repetition of errors and the speed at which users
dverted their gaze to the robot and the interface (Corrigan et al., 2014).
Synopsis

In most cases, an understanding of engagement is best approached through the combination of subjective and objective measures (LeeTiernan & Grudin, 2003; Stiubiener, Silveira, Matushima, Bressan, & Ruggiero, 2012). Those working with the concept are best served by a willingness to acknowledge the merits of a variety of approaches to assessment. Ultimately, the choice of measure depends upon our interpretation of engagement, epistemological leanings, need for methodological validity and reliability, the nature of the application context, and our motivations for measurement. Rendering these factors explicit enhances our capacity to advance measurement and understanding.

5.6 Design Strategies for Engagement

Engagement has been studied with respect to a highly diverse range of application contexts (See Table 5.4). Developers have adopted engagement metrics to improve learning outcomes in e-learning systems (H.-K. Chou, Lin, Woung, & Tsai, 2012), to support effective communication with personal agents (Beale & Creed, 2009), to interpret patterns of use online (Webster & Ahuja, 2006) and to understand the determinants of successful behaviour change interventions (T. Bickmore, Schulman, & Yin, 2010). Engagement has been employed with an eye towards not just analysis but synthesis; design. Strategies to support engagement have been studied with respect to museum exhibits (Dindler, Iversen, Smith, & Veerasawmy, 2010), mental health interventions (Doherty, Coyle, & Sharry, 2012), camera security systems (Shastri, Fujiki, Buffington, Tsiamyrtzis, & Pavlidis, 2010), games (Kurdyukova & Rehm, 2009), art (Edmonds, 2011) and more (Trotto & Hummels, 2013).

This section outlines strategies for the design of engaging experiences, as featured throughout this corpus. These strategies often reflect the relationship between conceptions of engagement with distinct disciplinary origins and a variety of unique design problems.

5.6.1 Ensuring Usability

Engaging experiences are built upon usable and functional forms of interaction. These systems satisfy the basic needs of use, are easy to use, reliable, learnable and result in the effective use of users’ time and resources (O’Brien & Toms, 2008).
### Table 5.4: Engagement & Application Context

<table>
<thead>
<tr>
<th>Application</th>
<th>Mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaming</td>
<td>56</td>
</tr>
<tr>
<td>General Learning</td>
<td>32</td>
</tr>
<tr>
<td>Online Learning</td>
<td>27</td>
</tr>
<tr>
<td>Social Media</td>
<td>15</td>
</tr>
<tr>
<td>Classroom Learning</td>
<td>14</td>
</tr>
<tr>
<td>Distance Learning</td>
<td>8</td>
</tr>
<tr>
<td>Virtual Learning Environments</td>
<td>8</td>
</tr>
<tr>
<td>Gamification</td>
<td>7</td>
</tr>
<tr>
<td>Virtual Agents</td>
<td>6</td>
</tr>
<tr>
<td>Algorithm Visualisation</td>
<td>6</td>
</tr>
<tr>
<td>Online Networks</td>
<td>6</td>
</tr>
<tr>
<td>Online Search</td>
<td>5</td>
</tr>
<tr>
<td>Online Services</td>
<td>5</td>
</tr>
<tr>
<td>General Education</td>
<td>5</td>
</tr>
<tr>
<td>Website Use</td>
<td>5</td>
</tr>
<tr>
<td>Museum Exhibits</td>
<td>5</td>
</tr>
<tr>
<td>Website Design</td>
<td>5</td>
</tr>
<tr>
<td>Social Networks</td>
<td>4</td>
</tr>
<tr>
<td>Robotic Agents</td>
<td>4</td>
</tr>
<tr>
<td>eGovernment</td>
<td>4</td>
</tr>
<tr>
<td>IVA Salesperson</td>
<td>3</td>
</tr>
<tr>
<td>Presentations</td>
<td>3</td>
</tr>
<tr>
<td>Reading</td>
<td>3</td>
</tr>
<tr>
<td>Robotic Tutors</td>
<td>3</td>
</tr>
<tr>
<td>Programming</td>
<td>3</td>
</tr>
<tr>
<td>Online Shopping</td>
<td>2</td>
</tr>
<tr>
<td>Storytelling</td>
<td>2</td>
</tr>
<tr>
<td>Exercise Support</td>
<td>2</td>
</tr>
<tr>
<td>General Computer Use</td>
<td>2</td>
</tr>
<tr>
<td>Robotic Companions</td>
<td>2</td>
</tr>
</tbody>
</table>

**Usability**

Although usability was studied infrequently within this sample of the engagement literature, one study of an e-learning system for hospital patients (n=281) found that “the user's intention to continue usage was predominantly determined by satisfaction with the system and perceived performance” (H.-K. Chou et al., 2012). A study of playability, usability and ‘long term’ engagement in games found that experiences can prove engaging despite usability defects if those problems are limited in frequency and the Density of Usability Defects (DUD) is low (Febretti & Garzotto, 2009).
Feedback

Prompt and accurate feedback supports engagement. One study of performance-related feedback reported that positive feedback motivated users to higher task performance (Fairclough, Gilleade, Ewing, & Roberts, 2013). Negative feedback did not lead to disengagement however, which the authors hypothesise could be due to the abstract nature of the task and a lack of negative consequences for failure. As a study of user attitudes towards online social experiences found, “pressure to perform” can be “both a good thing and a bad thing” (Social Computing Research Group, 2005). Another study manipulated the immediacy cues displayed by a human-like robot as it told stories (Szafir & Mutlu, 2012). Immediacy cues were defined as “actions taken by speakers to decrease the psychological distance between themselves and their listeners.” More adaptive, human-like behaviours led to improved recall, as well as greater rapport and motivation among female participants (Szafir & Mutlu, 2012).

Aesthetics

Given the ubiquity of high quality games and other consumer technologies, many users today may also hold aesthetic expectations which need to be met for engagement to ensue (H. van Vugt, Konijn, Hoorn, Keur, & Eliëns, 2007).

5.6.2 Managing User Resources

As a measure of interaction, engagement reflects the performance of both users and systems. This has led to its adoption as a means of ‘managing’ interaction, through real-time adaptation to the user. This strategy is most relevant to well-defined tasks with measurable outcomes. The measurable proxies for engagement in these contexts are often cognitive in nature, including difficulty (Castellano et al., 2009), cognitive load (Oviatt et al., 2008) and workload (Galán & Beal, 2012).

Challenge, Cognitive Load, and Workload

One study examined the implicit use of speech amplitude and pen pressure to engage a computer assistant while reducing users’ cognitive load (Oviatt et al., 2008). During a series of maths questions, 86% of intentional interactions were correctly identified through speech, and 75.2% through pen pressure, compared to human judgement. The reliability of the speech system improved over the course of the session as users adapted their behaviour to the system, often unconsciously according to post-session interviews.
Another system employed EEG measures of engagement and workload to predict whether 16 college students would provide the correct answers to a series of 8 maths problems (4 easy and 4 hard) (Galán & Beal, 2012). A prediction accuracy of 87% was achieved using the ‘most informative’ signals from both data streams. Workload scores were significantly higher for more difficult problems, while engagement scores showed no change, suggesting that engagement and workload could be treated as independent variables.

Another study found that playing games via a Brain-Computer Interface (BCI) introduced a higher workload than Automated Speech Recognition (ASR) control, although both systems ranked below the median of the Game Engagement Questionnaire (Gürkök, Hakvoort, & Poel, 2011). Other research has adopted workload as a direct proxy for engagement, including the development of a computational model to balance users’ task engagement (workload level) and automation (awareness level) (Klein & Van Lambalgen, 2011), and a 5-stage framework for the assessment of users’ cognitive engagement online, based on the concept of human mental workload (Longo, 2011).

Monitoring Engagement

When we speak of measuring engagement, often we speak not only of assessment for the purposes of research but of processes which could come to form a part of users’ daily lives. It is therefore essential to consider the acceptability of monitoring, and to respect users’ privacy and autonomy.

Systems designed to maintain user engagement often do so in order to optimise a particular outcome; “social interaction or efficiency of task completion” for example (Akker et al., 2009). However, increased engagement does not necessarily “imply improved efficiency in terms of task performance and vice versa” (Akker et al., 2009). Particular outcomes require appropriate forms of engagement. A system devised to facilitate public feedback in relation to public spaces, for example, was designed to engage users as “creative observers rather than mere information seekers” (Whittle et al., 2010).

Adaptive systems might not focus solely on the maintenance of engagement but also on signs of disengagement — discontinuities in interaction. Interruptibility has been examined with respect to task-focused systems, as in the case of an algorithm designed to interrupt programmers at the optimal point in time (Fogarty et al., 2005). Adaptation to user performance must be informed by sufficient context. If a user’s engagement with a game, for example, drops “because the player is bored, then the appropriate response from the system is to challenge the player.” If the decline in engagement is due to “excessive game difficulty,” this “requires a different adaptive response, e.g. to assist the player” (Fairclough et al., 2013).
5.6.3 Immersing Users

One strategy for engaging users focuses upon their immersion (deep psychological absorption) in experience.

**Immersion**

Immersion is often employed as a metonym of engagement, particularly in the contexts of gaming (Brown & Cairns, 2004; Carrigy et al., 2010) and virtual environments (Dow, 2007). As is the case for engagement, multiple theories and interpretations of immersion exist.

Immersion has been described as characterised by a sense of ease and familiarity, a more passive experience than engagement — a comparatively conscious and effortful process (Douglas & Hargadon, 2000). The same authors state that while flow hovers on a “continuum between immersion and engagement,” “immersion and engagement are neither mutually exclusive properties nor polar opposites.” The developers of a location-based mobile game also describe engagement as more active than immersion but ‘total immersion’ as an experience similar to flow (Carrigy et al., 2010).

In contrast, Swift et al. state that “a participant can be engaged in an activity without being immersed but cannot be immersed without being engaged” (Swift et al., 2010). This kind of hierarchical comparison features elsewhere, as in the statement that “it is not so much that involvement is necessary for learning, but that a certain amount of learning is required before being able to experience deeper levels of involvement” (Iacovides, 2009).

Another interpretation links immersion to the “sensation of being surrounded by a completely other reality … the experience of being transported to an elaborate simulated place” rather than to absorption in the present reality or task (See Murray, 1997 as cited in Carrigy et al., 2010). This distinction between immersion as “deep absorption” and “traversable space habitation” is respected by the Digital Game Experience Model (DGEM), which describes six *frames*, or modalities of meaning, which are gradually *internalised* by the player during the process of play (Calleja, 2007).

Other perspectives construe immersion as a psychological *state*, including Witmer’s view of immersion as “influenced by the immersive tendencies of the individual,” Salen and Zimmerman’s description of engagement as a process of double consciousness, and Ermi & Mayra’s SCI (sensory, challenge-based and imaginative) model of immersion (See Carrigy et al., 2010).

An alternative distinction has also been made between immersion, characterised by “loss of self-awareness, loss of social awareness, and loss of game awareness,” and what is termed *neo-immersion*, in-
volving “awareness of self, awareness of others, and awareness of the game” (Whitson, Eaket, Greenspan, Tran, & King, 2008).

Brown & Cairns define three levels of immersion, denoted engagement, engrossment and total immersion, drawing upon grounded theory analysis of interviews with 7 gamers following time spent playing their favourite game (Brown & Cairns, 2004). The authors describe the fleeting nature of total immersion as the distinguishing feature between it and the flow state. Brown & Cairn’s conception of immersion is linked to the idea that controls should be as close to invisible as possible (Turner, 2010; Jacucci et al., 2010). This model provides a strategy for the design of engaging experiences by defining the barriers between multiple ‘depths of interaction.’ Barriers of access (related to the game controls and feedback), time, effort (investment, expectation of rewards) and attention (a willingness to concentrate) pertain to the first state, game construction (visuals, interesting tasks, plot) to the second, and empathy and atmosphere (barriers to presence) to the third (Brown & Cairns, 2004).

Presence

Presence is closely linked to immersion, and to the hypothesis that greater invisibility of controls leads to greater immersion (Brown & Cairns, 2004). This concept is often encountered in the context of augmented or virtual reality. Biocca’s three forms of presence, physical, social and self, suggest however that this term cannot be tied to a single conception (See Dow, Mehta, Harmon, MacIntyre, & Mateas, 2007). Presence has also been described as perceptual immersion but not psychological immersion (Seah & Cairns, 2008).

Dow also views presence as ‘perceptual immersion,’ distinct from engagement; “a user’s involvement or interest in the content independent of the medium” (Dow, 2007), and further states that while the aim of virtual and augmented reality systems is to make users feel “as if they are ‘there’” or to “create the ‘illusion of nonmediation,’ this can fail “to recognize the notion of actually engaging the user” (Dow, 2007). This claim is supported by a study of three unique implementations of the interactive drama Façade — using text-input, speech and augmented reality (AR) (Dow et al., 2007). Only half of players (6 of 12) reported preferring the AR version of the system, citing a desire for more distance from the game in order to engage more comfortably with it. This suggests that “more ‘natural’ or ‘transparent’ interfaces, where the medium fades to the background, will not necessarily lead to a compelling game or entertainment experience” (Dow, 2007).
Involvement

Involvement, a similar concept to immersion, has also been compared to engagement (Kappelman, 1995). It has been claimed that engagement implies a degree of intrinsic interest, whereas involvement may occur even when the user experiences no enjoyment (Webster & Ahuja, 2006).

5.6.4 Engaging Users Emotionally

There are strong parallels between engagement and affective computing research, given the motivation to produce interactive systems with representations of fuzzy states (Beale & Creed, 2009). Engagement itself has also been described as an affective state or emotion, and the measurement of engagement as an emotional response has been used to support the design of engaging experiences.

Defining Affect

Affect is most commonly described as the experience of emotion, comprising dimensions of valence and arousal (D'Mello et al., 2010). The 9x9 affect and arousal grid, for example, has been directly applied to the measurement of engagement, in the context of collaborative lecturing by human and computer tutors (ibid.). An alternative view positions engagement (as attention) as a moderator of affect which filters perception. One study of art viewing found that “momentary engagement overrides the effect of mood on ambient/focal attention” (Biele, Kopacz, & Krejtz, 2013).

Measuring Affect & Engagement

The measurement of affect and engagement share many of the same challenges and techniques. In the realm of affective cinema, for example, engagement has been assessed as ‘positive excitation’ according to Galvanic Skin Response (GSR), EEG and automatic facial tracking measures (Abadi et al., 2013). In the context of two insightful studies of audience engagement, engagement was described as “related to attention and interest” although operationalised as valence and arousal (Latulipe et al., 2011). During the first study, 3 choreographers and 4 theatre directors were shown video of a performance (dance or theatre) alongside a time-line of previous viewers’ galvanic skin responses (reflecting arousal) (Latulipe et al., 2011). This software allowed users to vary the time period across which an average GSR response was calculated. This practice highlighted the need to differentiate between semantic-chunking and stimulus-response interpretations of data. The narrative structure of a live performance means that
interpreting momentary data as though at “each second, audience members are responding to the thing they just saw happen in the previous second” is clearly erroneous, as stated expressly by participants.

A second study examined audience reports of engagement. Nine participants watched a video of an 11 minute dance performance in an immersive environment while wearing a GSR sensor and manipulating a slider to indicate their level of engagement. Multiple ‘vocabularies of engagement’ were tested. With respect to a scale comprising ‘No Engagement to High Engagement,’ it was found that users “could not detach valence from the word, and tended to only rate themselves as being engaged when they liked what they saw. Others just didn’t really seem to know what we meant by engagement, and still others didn’t seem to know how engaged they were” (Latulipe et al., 2011). Other vocabularies studied included ‘Love it! to Hate it!’ and ‘No Emotional Reaction to Strong Emotional Reaction.’

When later presented with a graph of their responses and asked to explain their ratings, those in the emotional reaction group did not provide as many details about the performance and several reported difficulty using the scale. Those in the Love-Hate group could “lower their love-hate rating if something confused them” but “the mapping to the ER scale” was “less clear.” This was the only work within this corpus which examined users’ interpretation of engagement despite the potential significance of these differences, particularly in the context of self-report methodologies (See Chapter 3).

5.6.5 Inspiring Fun

Engagement has been conceived not only as utilitarian, but also as driven by enjoyment, gamification, escapism and fun; “the state of mind that we must attain in order to enjoy a representation of an action,” entailing “a kind of playfulness — that ability to fool around, to spin out ‘what if’ scenarios.” (Laurel, 1991; Webster & Ho, 1997). Ludic engagement has been defined as “a precondition to play” (Brooks, 2013). Enjoyment has been described as a component of engagement (Rozendaal et al., 2008), and characterised as composed of multiple constructs, including engagement, positive affect and fulfilment (See Warner’s model as cited in Lin, Fernandez, & Gregor, 2012).

Humour

Humour has also been shown to support engagement. The addition of a sidekick robot to a child-robot interaction, for example, increased the attention users paid to the spoken dialogue and led twice as many participants to laugh at least once (Vázquez, Steinfeld, Hudson, & Forlizzi, 2014). Analysis conducted in the context of video games has concluded that “humour itself stimulates engagement and provides a pleasurable experience” (Dormann, Barr, & Biddle, 2006).
Gamification

Gamification features prominently among approaches to the design of enjoyable experiences. These methods often attempt to leverage economic and social psychological motivations, including social facilitation and social loafing effects (Liu et al., 2011). Although economic incentives may increase the quantity of ‘work’ completed, they cannot be guaranteed to increase its quality (Liu et al., 2011). These effects can be transitory, “participants can feel manipulated,” “little or no transfer can occur if behavior is only driven by rewards” and “if the reward vanishes so does the behavior” (Domínguez et al., 2013).

During one attempt to ‘gamify’ an e-learning experience, the authors’ use of competition as an incentive was not well-received by all students, one participant stating that “it would be more interesting to improve the traditional version, instead of making competitions” (Domínguez et al., 2013).

A study of children’s engagement with LIEKSMYST, a hypercontextualised game (HCG) in a museum environment revealed four motivational factors; curiosity, challenge, fantasy and personal control (Islas Sedano et al., 2013). The authors stress the importance of freedom and voluntary participation for engagement; “If a game has to be played as part of a prescribed school assignment, then the pupils (or players) do not access that game voluntarily. Instead the game is viewed as a task that the pupils have to complete within the school milieu.”

It is also important to note that negative experiences can serve as catalysts for enjoyment and engagement. Users’ time spent playing games, for example, is often strongly motivated by negative reasons for playing (Boyle, Connolly, Hainey, & Boyle, 2012).

5.6.6 Enabling Exploration

The use of any product or service encompasses more than the pursuit of a single, isolated goal, but the engagement of additional senses, actions and feelings (C.-J. Chou & Conley, 2011). Engaged reading, for example, has been described as “not a mere information-extracting process rather an experience encompassing a wider range of activities” including “book selection, reading, imagination, learning, multisensory exploration, confrontation with the peers, etc.” (Colombo & Landoni, 2011). Tasks have been classified as experiential or goal-directed (Rozendaal, Keyson, & de Ridder, 2007; Rozendaal et al., 2008) and ill or well-defined (Goldberg et al., 2011; Brooks, 2013). Experiential, or ill-defined, tasks can result in richer experiences, allow for exploration and prove more engaging (Brooks, 2013; Kucirkova et al., 2014).
Richness

The RC & E framework models engagement according to the experience of richness and control. Within this framework, richness is defined as the “complexity of thoughts, actions and perceptions … evoked during the activity” and control as the “effort that is experienced in the selection and attainment of goals” (Rozendaal et al., 2008). Engagement is defined as the square root of the product of richness and control. In a study of user engagement with a digital table-top game, richness was controlled by modifying the variety of ball and user behaviours, level of colour, detail and asymmetry of the visual design, and control by manipulating the hand-eye coordination required and by the addition of redundant visual information (Rozendaal et al., 2007). It was found that richness commanded more of an impact on engagement during experiential tasks while control had a greater effect during goal-oriented tasks. When presented with a well-defined task, users appear to prioritise the ease with which this can then be accomplished. A lack of richness can lead to boredom and a lack of control to anxiety (Rozendaal, Keyson, Ridder, & Craig, 2009).

The study of narrative is also related to the richness of experience. In the context of narrative multimedia, engagement is cast as supported by spatial containment, causality, skill-based interaction, dialogue causality, invisibility of the medium and the illusion of intelligence (Mallon & Webb, 2000). A potential conflict exists between the provision of control to the user and disruption to the narrative flow however. Adams writes that although “interactivity and storytelling” are not necessarily mutually exclusive, “the more you have of one, the less you're going to have of the other” (E. Adams, 1999).

Novelty

Exploration requires novelty. Flow theory views novelty not only as related to first use but as a dynamic component of interaction (Cowley et al., 2008). Together, engagement and novelty may inform a user's decision to participate in an experience (Cowley et al., 2008). However, novelty has only rarely been examined with respect to engagement (Webster & Ho, 1997; Bianchi-Berthouze et al., 2007; O’Brien & Toms, 2010a).

5.6.7 Supporting Social Connectedness

Some of our strongest motivations for engagement with technology are social in nature. Social engagement has been described as “the first stage of social intercourse” (Singletary & Starner, 2001), and also as “not just about how a single interaction unfolds, but about how and why people develop a relation-
ship with a platform or service and integrate it into their lives” (Jaimes et al., 2011a; Jaimes, Lalmas, & Volkovich, 2011b). The link between social interaction and engagement has been studied within sociology (Goffman, 1978), but also with respect to technologies for gaming (Lindley, Le Couteur, & Berthouze, 2008; Rozendaal et al., 2009; Abreu, Gho, & Byrne, 2013; Oksanen, 2013), learning (McClure, 2013), social networks (Freyne, Jacovi, Guy, & Geyer, 2009; Kirman et al., 2010), and joint media engagement (Ballagas et al., 2013).

Social Strategies

The strongest motivations for participating in social networks have been found to be associated with a sense of belonging and influence, specifically efficacy, anticipated reciprocity, and increased reputation and recognition online (Freyne et al., 2009). One study found that the users of a social network who received people recommendations were more likely to return to the site. Those who received recommendations to active people were most likely to return (Freyne et al., 2009).

*Mutual engagement* has also been studied with respect to Daisyfield, a collaborative music making program which allows users to create musical loops in a shared visual space (Bryan-Kinns, 2013). Participants were most mutually engaged when granted either the provision of identity cues or indicators of each other’s focus of attention, but not both. The author suggests that “maybe there was an expectation that others would take more notice of their work when they could see each others’ loci of attention,” leading to “increased expectations on others — a form of social contract.”

Another study of collaborative music making identified three ‘qualitatively different engagement relationships,’ individual, unilateral, and bilateral (Swift et al., 2010). Participants commented that they viewed individual engagement as a potential barrier to group enjoyment. The use of *network theory* and real-time visualisation of these relationships was therefore proposed as a means to support engagement.

A novel photo apparatus, ‘Playful Booth’ was designed to engage children in social interaction in order to facilitate analysis of their social behaviours (Lyra, Karapanos, Gouveia, Nisi, & Nunes, 2013). Children reacted positively to the booth, engaged by ‘the magic of the RFID watch functionality,’ a sense of playfulness, ownership of their watches, and the act of photo taking in a public space, which induced a sense of belonging, friendship, public performance, and exclusivity.

Social Presence

The link between social presence and engagement has also been studied, including with respect to education (Wise, Chang, Duffy, & del Valle, 2004) and human-robot interaction (J. J. Choi et al., 2014).
Biocca et al., for example, describe social presence as composed of three dimensions; co-presence, psychological involvement and behavioural engagement (See Martino, Baù, Spagnolli, & Gamberini, 2009). Users have been described as socially present when “interdependent on other people inhabiting the same environment” or when they “intensely feel” engaged with other agents (Martino et al., 2009), underlining the embodied nature of engagement.

5.6.8 Enabling Optimal Experience

Engagement has also been adopted as a means to assess progress towards a state of ‘optimal experience’ — a target for designers. The prominent adoption of flow theory frequently reflects this perspective (Cowley et al., 2008).

The Flow State

States of flow and engagement are often viewed as synonymous. Engagement has also been described as a subset of flow and a more passive state which is therefore more appropriate than flow when the user experiences less control (Webster & Ahuja, 2006). It has also been suggested that games which do not set clear goals, as required for flow, can still prove immersive, and that negative, and even frustrating, experiences do not necessarily lead to decreased immersion (Seah & Cairns, 2008).

Systems which adapt to the user have been explored as a means of facilitating the flow state. This includes adaptive algorithms which determine when a player should be allowed to progress during a learning game (Lomas, Stamper, Muller, Patel, & Koedinger, 2012) and a model which adapts to the gameplayer, not by adapting the level of challenge, but by changing the way the player perceives the results of his or her actions in order to improve their confidence (Van Der Spek, 2012).

Flow theory is typically employed with respect to brief time scales, and has likely contributed to the prominence of situated experimentation within the engagement literature. However, it has also been suggested that flow theory could be applied to interaction “over a much longer time scale” (Seddon et al., 2008).

5.6.9 Promoting Repeated Use

Users come to interact with products and services through a combination of personal volition and environmental factors. However, many systems require active rather than passive users (Webster & Ho, 1997; Karavirta & Korhonen, 2006; Fischer & Benford, 2009) and engagement “both in the moment and over time” (Lamberty, 2004). This is typically the case for EMA technologies.
One strategy for engagement in this context is to reduce the burden on users by designing for simplicity and efficiency (Scollon et al., 2009). The EMA literature calls for the “brevity and simplicity of the user interface and input” (Koop, 2002), “a very simple, clean, and easy to use platform” (Lanzola et al., 2014) where “the number of possible actions should be kept to a minimum” (Palmblad & Tiplady, 2004). Users’ comments tend to support these goals; “it’s simple and is actually uncomplicated. There’s not much interface necessary” (Möller et al., 2013).

A more proactive approach is to employ extrinsic incentives, such as monetary remuneration which might vary by rate or frequency and increase over time (Christensen et al., 2003). Presenting participants with a running reward total can encourage further participation (Aharony et al., 2011). However, too high a reward can lead to poor data quality, due to selection biases for example (Scollon et al., 2009). One intensive study of cigarette smoking among adolescents employed monetary incentives (Mermelstein et al., 2007). The authors write that paying participants “$40 for the baseline EMA weeklong, with escalating payments over waves” led participants to treat the task as a job, and “one they had to take seriously.” The researchers also worked to create “a collaborative research relationship,” “through one-on-one in-person training, personalized follow-up phone calls, and consistency in contact people,” which “was important to the success of the data collection” (Mermelstein et al., 2007). Emphasising the importance of participants’ contribution to science requires establishing a ‘viable research alliance’ (Christensen et al., 2003; Scollon et al., 2009). Booster telephone calls and reminder emails have also been used to induce compliance (D. J. Beal & Weiss, 2003; Möller et al., 2013). One study found that SMS reminders sent 10 minutes following a missed prompt led to a “10% increase in response rates” (Hofmann & Patel, 2015).

Intrinsic incentives can also motivate a habit of use. An EMA study of mood, interruptibility and computer use found that providing participants visualisations of their data resulted in a 23% higher compliance rate (Hsieh et al., 2008). UbiGreen, a mobile application for tracking transportation habits, employed an evolving depiction of an arctic scene or tree to reflect the environmental impact of users’ transport behaviours (Froehlich et al., 2009). A similar approach has been used to encourage physical activity (Consolvo et al., 2008). Presenting response rates has alone been shown to boost protocol compliance (Christensen et al., 2003). Providing feedback has also been found to encourage use by conveying the impression that someone is “closely monitoring … progress” and “cared about the information” (Koop, 2002). While providing feedback to users may motivate longitudinal reporting (Ramanathan et al., 2012), this approach may also bias reports, inspire increased reflection and lead participants to “rate their present state in reference to their previous states” (Scollon et al., 2009).
Designing for engagement with respect to self-report technologies therefore typically entails reducing the burden on users by implementing a simple and efficient reporting process while at the same time providing intrinsic and extrinsic incentives for use, and attending to the challenges of reactivity, habituation, validity and data completeness (Koop, 2002; Christensen et al., 2003; Palmblad & Tiplady, 2004; Mermelstein et al., 2007; Ebner-Priemer & Trull, 2009b; Lanzola et al., 2014). The merit of a variety of incentives, including monetary remuneration, booster telephone calls, reminder emails, data visualisation and feedback, remains an active topic of research (Rabbi et al., 2017).

Often overlooked in the pursuit of engagement are the potential consequences of its excess.

Addiction

While negative patterns of engagement can take many forms, game addiction has received the most significant focus. Several scales have been developed to assess addiction and its relationship to engagement, including Brown’s Addiction Checklist, Charlton’s Computer Apathy and Anxiety Scale (CAAS), Tellegen’s Absorption scale (Seah & Cairns, 2008), and the Addiction-Engagement Questionnaire (Metcalf & Pammer, 2011). An attentional bias for gaming-related words has also been identified among addicted online game players (Metcalf & Pammer, 2011). Addiction and engagement have been described as “attitudes to the general experience of playing,” and immersion as closer to “the actual experience of specific instances of play” (Seah & Cairns, 2008).

One study examined gaming addiction among 1,332 South Korean middle and high-school students using a 65-item questionnaire (Seok & DaCosta, 2012). Between 1.7% and 25.5% of the sample group were classified as addicted depending upon the criteria and cut-off threshold adopted. The authors stress that further work is needed to reach a consensus on criteria for distinguishing between non-pathological and pathological gameplay, if reliable prevalence rates are to be obtained.

Many technologies are now designed to attract and maintain user engagement, including mobile games, social networks and news applications. These systems provide content which is personally tailored, packaged into small chunks, updates frequently and unpredictably, and whose delivery is activated at the press of a button. There are clear parallels between these strategies for engagement and those employed in the design of the slot machines of Las Vegas (Schüll, 2012).
5.7 Discussion

The findings of this review support further reflection on the past, present and future adoption of engagement.

5.7.1 Domains of Knowledge

Interpretations of engagement extend beyond what we might think of primarily as ‘user engagement.’ Divergent domains of knowledge can therefore provide unique forms of insight.

Conversational Engagement

Forward-looking work on engagement is to be found in the study of human-computer conversation, including the development of robots and virtual agents (C. Sidner & Dzikovska, 2002; C. L. Sidner et al., 2004, 2005; Baecker et al., 2006; Rich, Ponsleur, Holroyd, & Sidner, 2010; Xu et al., 2013; Fagen & Kamin, 2013; Oertel, 2013; Oertel & Salvi, 2013). In this domain, interaction is often interpreted though conversational analysis. This perspective features a unique distinction between backward and forward-looking forms of engagement, the first referring to grounding behaviours, “what is understood up to the present utterance,” and the latter to users’ need to “be connected and aware in the interaction” (C. L. Sidner et al., 2005).

Information Visualisation Engagement

With respect to information visualisation, engagement has been explored as a means of balancing design goals of cognitive efficiency and learning, “a trade-off between efficiency and beneficial obstructions” which promote active processing (Hullman, Adar, & Shah, 2011). The ITiCSE Engagement Taxonomy characterises engagement as a ‘type’ or ‘depth’ of interaction comprising six forms; No Viewing, Viewing, Responding, Changing, Constructing and Presenting (Naps et al., 2002). This taxonomy has been applied to several design scenarios (Karavirta & Korhonen, 2006; Urquiza-Fuentes & Velázquez-Iturbide, 2012) and attempts made to extend it (Mylær, Bednarik, Sutinen, & Ben-Ari, 2009).

Learning Engagement

Definitions of learning engagement are often brief (Grimes, Warschauer, Hutchinson, & Kuester, 2005; Cocea & Weibelzahl, 2007; Barkatsas, Kasimatis, & Gialamas, 2009; Hu & Hui, 2012; McCay-Peet et al., 2012; Blasco-Arcas, Buil, Hernández-Ortega, & Sese, 2013). It has been argued in this context
that research often “essentializes engagement,” ignoring the effects of gender, socio-cultural, ethnic and economic status (Zyngier, 2007, 2008). School engagement has been described as malleable, responsive to contextual features, amenable to environmental change, and as a ‘meta’ construct (Fredricks, Blumenfeld, & Paris, 2004). Engagement has been construed by practicing educators as a property of materials, as learner attentiveness, as interaction and as increasingly legitimate participation within a community of practice (Notess, 2008). With respect to this last conception, engagement is “the result of learning as much as the means of learning.”

Gamification is often employed as a strategy to support student engagement. Although there have been attempts to guide designers (Schuurink et al., 2008), this approach has not always met with success. A systematic review of computer game and serious game use found no strong evidence to suggest that games lead to more effective learning (Connolly et al., 2012). A weakness of games designed for learning may be that they do not sufficiently inspire reflection. Avoiding "opportunities that users might take to reflect on their actions" has even been described as a trick used by designers to maintain player engagement (Whitson et al., 2008). It could be that slower forms of interaction are required to promote complex cognitive processes such as understanding (Kuznetsov, Harrigan-Anderson, Faste, Hudson, & Paulos, 2013) (See Chapter 3).

Civic Engagement

Civic engagement is a broad field of research, encompassing e-government, citizen sensing projects, gendered engagement in STEM subjects, community-based activism and the generation of social capital (B. J. Kim & Kleinschmit, 2012). Definitions of civic engagement have been proposed by a number of authors, although it is often in a transitory fashion that references to user engagement with technology are encountered within this domain (Park, 2006; Farnham, Keyes, Yuki, & Tugwell, 2012; Taylor et al., 2012; Nepal, Sherchan, & Paris, 2011).

Wellbeing Engagement

Engagement has also been studied with respect to wellbeing (See Chapter 2), including, for example, an intervention designed to boost social engagement between residents of retirement homes (Linnemeier, Lin, Laput, & Vijjapurapu, 2012) and a serious game devised to reduce work-related stress while increasing work-related engagement (Wiezer, Bakhuyse Roozeboom, & Oprins, 2013). One interpretation of engagement in this context is “a kind of mindfulness requiring cognitive effort and deep processing of new information” (Xie, Antle, & Motamedi, 2008). However, it has been suggested that
although “characteristics such as feedback, control and interactivity may be associated with systems that are likely to promote engagement, they may not necessarily promote the mindful kind” (J. Pearce & Pardo, 2008). Studying user engagement can provide insight into the role we wish technology to play in our lives, possibly supporting a shift from “proactive computing to proactive people; where UbiComp technologies are designed not to do things for people but to engage them more actively in what they currently do. Rather than calm living it promotes engaged living” (Rogers, 2006).

5.7.2 What’s Missing?

A number of gaps are to be found within the engagement literature.

Engagement in The Real World

Most studies of engagement take place in constrained laboratory environments, establishing internal validity but often neglecting ecological validity. There exists a need for research in real-world contexts; “open, relatively unconstrained environments, where multiple people with different needs, goals and long-term plans may enter, interact and leave the observable world” (Bohus & Horvitz, 2009a, 2009b, 2009c). The design of virtual agents, robot hosts, mobile applications and systems of all kinds rests upon an understanding of experience in daily life.

Voluntary Engagement

“Providing participants with instructions, such as ‘Go interact with this system’, or ‘Go join the existing interaction’ can significantly prime and alter the engagement behaviors they would otherwise display.” Bohus & Horvitz, 2009c

The most significant gap in the literature however concerns voluntary engagement. Supporting voluntary participation in user studies is challenging but an essential component of engagement research (Bohus & Horvitz, 2009c; Islas Sedano et al., 2013). A focus on voluntary use is needed to shift design towards more ‘realistic’ forms of engagement, as expressed in the framing of the focused-casual continuum for example (Pohl & Murray-Smith, 2013). Rather than focusing primarily on achieving intense degrees of engagement, designers might permit users to choose appropriate ‘levels of interaction,’ including more casual forms, “depending on the situation” (Pohl & Murray-Smith, 2013). Novel methods of recruitment, such as Google AdWords (Colbert & Boodoo, 2011) or mobile applications released for public use (Killingsworth & Gilbert, 2010) merit further exploration. It is logical that the study of engagement should allow for realistic means of engagement, and yet this is in fact rare.
Engagement Over Time

Another prominent gap within the literature pertains to the study of engagement over time. Definitions of long-term engagement include the “maintenance of user adherence to a desired interaction usage pattern” (T. W. Bickmore, Consolvo, & Intille, 2009) and “the degree of voluntary use of a system along a wide period of time (i.e., weeks, months, or years), involving dozens, if not thousands, of interactions, each one spanning for significantly longer than [sic] few seconds or minutes” (Febretti & Garzotto, 2009).

Longitudinal engagement effectively has a vocabulary all of its own; patterns, trajectories, sporadic, immediate, periodic, shaped, sustainable. These terms are empowering in that they tie often verbose conceptions of engagement to a real-world context. Experience is cumulative (Carrigy et al., 2010), motivations for use evolve over time (Naps et al., 2002; H. C. van Vugt, Hoorn, Konijn, & de Bie Dimitriadou, 2006; Snow, Buys, Roe, & Brereton, 2013), and so engagement cannot be fully understood without longitudinal analysis (M. Chen et al., 2011).

Long-term research has produced insightful results. A meta-study found that “students’ utilization patterns of visualizations had a much greater impact on their learning success - and therefore on the tool’s educational effectiveness - than the quality of the visualizations” (See Hundhausen, Douglas, & Stasko, 2002 as cited in Isohanni & Knobelsdorf, 2011). Another study found that increased engagement with virtual agents designed to support behaviour change in fact led to a decrease in exercise over time (T. Bickmore et al., 2010).

However, a call for increased analysis of longitudinal engagement does not imply that continuous or consistent engagement is always the appropriate design goal. Also lacking within the literature is discussion of the processes of disengagement, dynamic and stepped models of engagement, and non-use. Longitudinal research is currently overshadowed by a multitude of brief and situated experiments.

Engagement With Design

Another important, and often overlooked, feature of engagement research is the fit between various conceptions of engagement and the design process. The design of engaging experiences hinges upon the assessment of user engagement. However, little attention has been paid to the integration of measures of engagement within a process of design.

While several rigorous user engagement questionnaires have been developed, it is not clear how useful these measures have been to practicing designers. Of interest would be an exploration of how
the UES (O’Brien & Toms, 2010a) might fit within an iterative design process, in comparison to the ESQ (Schoenau-Fog, 2011) for example. Depth of analysis, efficiency of application and ease of interpretation are all desirable qualities for such measures. Would a One Minute Questionnaire (King & Robinson, 2009) provide actionable insight for engagement? Is asking users ‘Was this an engaging experience? If not/so, why?’ as useful a means of evaluation as validated questionnaires, interviews, behaviour logging or observation with respect to particular design goals? What is clear is that there is room for further analysis of the value of different methods of assessment. Many of the barriers to the adoption of engagement are familiar to designers, including the challenge of highlighting “relevant factors of engagement while still allowing designers to use their intuition and creativity” (Rozendaal et al., 2008) and the lack of optimal solutions to design problems (Hartevedt, Thij, & Copier, 2011) (See Chapter 1).

One of the strongest motivations for a study of engagement lies in its visceral nature, the fact that it is so intimately tied to our lived experience. However, the gaps highlighted here represent almost the antithesis of this ethos — a longitudinal outlook, voluntary use, both micro and macro-theoretical perspectives (See Table 5.2), and a focus on real-world environments. If engagement is to act as the foundation for a useful study of human behaviour, the most important feature of future work will be itself an engagement with the reality of lived experience.

5.7.3 Understanding Engagement

Understanding engagement requires embracing a diverse methodological toolkit in light of the merits and limitations of each perspective. These challenges have been encountered elsewhere within HCI, with respect to user satisfaction (Lindgaard & Dudek, 2003), usability (Hornbæk, 2006) and emotion (Hudlicka, 2008) for example. In this section of our discussion we propose several means of negotiating the complexity of these concepts.

Strive for Clarity of Expression

Engagement is a term we use to share meaning with respect to certain qualities of experience. However, language is flexible and interpretation reliant upon cultural and professional bias. It is unhelpful to find ourselves caught up in questions such as whether the statement that “a student is engaged by a programming project, but is interested in computer science” (Hansen & Eddy, 2007) is ‘true.’ Or whether a related concept such as ‘interest’ is a feature of engagement, an outcome or a contributing factor. Or between antonymic (engagement v. disengagement) and valenced (positive v. negative) interpretations.
A scaffolding is required to support the expression of meaningful differences between conceptions. This can be provided by awareness and explicit expression of ontological and epistemological positions, levels of abstraction and analysis, intentionality, the framing of an HCI ecology (See Section 5.4), theoretical foundations (See Section 5.3) and motivations for the adoption of engagement (See Section 5.5). Ultimately, we must strive for clarity of expression in interaction with researchers, designers and users.

**Embrace Process**

Understanding engagement is a process in which researchers, designers and users all have a part to play. Engagement refers not only to the use of technologies but also to involvement and participation in their design (Hwang & Thorn, 1999).

**Explore Hidden Assumptions**

The engagement literature features many ontological and epistemological perspectives, from post-positivist attempts to implement models of engagement to interpretivist analyses of ethnographic data. Clearly articulating what we mean by engagement necessitates an understanding of our own assumptions with respect to interpretation of the concept. The *structure and agency problem*, for example, questions whether it is “the social context in which individuals act that guides, determines or facilitates their actions or whether it is the individuals (or actors) themselves who form and shape the social context and institutions around them” (Grix, 2010). The problem of *reflexivity* calls for a focus on the ease with which it is possible to move “between concrete data and abstraction and back again” (Grix, 2010).

**Learn from Parallel Trends**

We discern a similar trend between work on engagement and related concepts such as experience and emotion. We refer to the call for a ‘turn to practice’ with respect to *user experience* — a turn away from the dominant perspective of rationalism, which can present “an obstacle to thinking about technology by reifying technological artefacts as objects of study apart from their making and use” (McCarthy & Wright, 2004). A similarly prominent rationalist perspective is often found with respect to the study of *affective computing*, for example, where researchers work to model emotions as discretised states, a “transformation from signal to symbol,” tackling emotion as a “problem of representation” best addressed by state-based models (Picard, 2000). Likewise, there has been a call to look to the ‘social-interactionist’ alternative, to acknowledge the social and cultural origins of emotion and its interpretation as experience, not only as an objective, measurable unit (Boehner et al., 2005).
The user engagement perspective has also proved predominately rationalist to date, as evidenced by the adoption of both theory (See Table 5.2) and methodology (See Table 5.3). This is not necessarily an unfavourable finding. However, we would echo, with respect to engagement, the call for “an appreciation not only of how emotion as a phenomenon is measured but also how measures make the phenomenon of emotion” (Boehner et al., 2005). It is worth widening our analytical focus in search of a more complete understanding of engagement.

Embrace Methodological Diversity

Certain products and services aim to realise a greater range of experience, emotional, cognitive, behavioural and social, than do others. This inevitably leads to conceptions of engagement and experience which differ greatly with respect to their complexity, and therefore necessitate diverse means of evaluation. The entanglement of engagement, experience and their component parts however, need not prove overwhelming, given appropriate clarity of expression. We often employ resting heart-rate as a useful measure of health and wellbeing without any confusion concerning its limitations. The implications of any one measure of engagement for a user’s experience must be interpreted in similar terms.

Don’t Confuse Models with Reality

We must take care when creating models of engagement to acknowledge their limitations. Many approaches we deem rational and objective rely on human judgement to establish ground truth, often through self-report or annotation. Even as humans we can only make inferences about another’s state of engagement. It is therefore important to avoid committing a ‘fallacy of reification,’ mistaking a model of engagement for ‘the concept itself.’ A map is not the territory it represents (Korzybski, 1933).

Focus on Fit for Purpose

Certain definitions of engagement infer a direct link between conception and assessment. It has been argued, for example, that “[i]f engagement to you is repeat visitors [sic] by visitors then call it Visit Frequency, don’t call it engagement” (Kaushik, 2007). More complex conceptions of engagement may require significant inference. However, acknowledging the conceptual nature of engagement must not be interpreted to legitimise all possible approaches to assessment and interpretation. The established metrics of validity, reliability and generalisability maintain their significance. It is also important to note that once users become aware of the adoption of engagement as an outcome measure, this metric can cease to reflect behaviour accurately (Goodhart, 1975).
Similarly, when designing for engagement, it is essential not to overlook other more pertinent or important objectives. Any one interpretation of engagement is unlikely to prove fit for all purposes and greater engagement does not necessarily imply enhanced learning, wellbeing or experience (T. Bickmore et al., 2010; Connolly et al., 2012).

Acknowledge Complexity

Engagement is a ‘device’ we adopt to facilitate a particular perspective on interaction and experience. It does not exist predefined in nature, nor is there one universally ‘optimal’ approach to its understanding. We are best served by asking not ‘what is engagement?’ but ‘which interpretation of engagement is best suited to this context and purpose?’ while justifying the choices we make — moving on from unhelpful and unfounded discussion of what engagement ‘is’ towards how engagement is best ‘interpreted’ and ‘measured.’

5.7.4 The Value of Engagement

Engagement is undoubtedly a popular concept. However, the literature has featured very little discussion of the value it is perceived to provide. Here, we explore the value of the concept as implied by its conception, measurement and adoption to date.

Value to Research

Engagement is frequently adopted not just an outcome in itself but as conducive of other aims. It is both an end and a means to an end. Engagement is popular in educational research, for example, due to the assumption that greater engagement with a system for e-learning is likely to support improved learning outcomes, and in wellbeing research, behavioural outcomes. Engagement, from this perspective, is seen as part of a larger whole, a component of a more complex abstraction, such as wellbeing, happiness, learning, productivity or task efficacy. This mediating relationship is one of the potential advantages of engagement over the concept of experience. It supports a reasoned connection to other valued concepts.

Engagement describes and supports collaboration, involvement, perseverance, attention, participation, attachment, closeness, adherence and more. Engagement is viewed as a flexible concept. This can contribute to ambiguity but renders engagement relevant to a variety of contexts and purposes. It also supports multiple levels of analysis. Engagement can be viewed as ephemeral and visceral in nature but can also be linked to, and have important consequences for, more abstract cognitive processes,
including not just learning but reflection (Whitson et al., 2008), empathy (Hall, Woods, Aylett, Newall, & Paiva, 2005) and understanding (Davies, 2002). This means that engagement can provide insight into user experience at multiple depths, spanning the gap between theory and practice. This turns our attention to a spectrum of use, degrees of engagement and disengagement, as well as change over time.

Value to Design and Development

One perspective on engagement might view it as the process by which we frame experience. This provides a grammar for discussing patterns of engagement and differences between users — a grammar which can help inform design.

Evidence of voluntary engagement with a system indicates that the system is not just usable but useful. It provides value, in some form, to its user. User engagement can therefore be viewed as a measure of the ‘health’ of an interactive system, of its value, as inferred by proxy of its measurable components, not blood pressure and heart rate, but interaction and subjective experience. Engagement is seen to imbue many of the positive features of experience and to present easily measurable forms. However, there are limits to such analysis. Engagement is most insightful when use of a system is voluntary and takes place over time. It is not always positive in its outcomes, motivations or design, as evidenced by technologically-driven distraction, social withdrawal and addiction.

Knowledge of user engagement allows designers to grasp the value of their products to the user, and sufficient detail can provide actionable insight to support the iteration of design. Whether engagement, and of what kind, is the appropriate design goal is a determination which needs to be made. However, the concept offers a pathway for action, a target, in a way which the comparatively amorphous concept of experience may not.

Value To The User

Engagement represents the focus of one’s conscious experience. The redistribution of this capital is one of the key roles technology plays in our lives. However, engagement can also be viewed as empowering, emphasising users’ autonomy and control. A subtle difference between ‘use’ and ‘engagement’ can be discerned for example. Engagement tends to emphasise a user’s autonomy, needs, motivations and the promise of a positive experience. Use can appear disembodied in comparison. A focus on engagement presents technology as ‘used’ by humans to varying degrees of intensity, encouraging us to take seriously the role we wish technology to play in our lives.
5.8 Conclusion | Reflections & Takeaways

To study engagement in terms as broad as those of this chapter is essentially to study the pursuit of knowledge — through a unique conceptual lens. This is an epistemological choice which reflects a belief in the capacity of this framing to support understanding of a form conducive to the aims and aspirations of design research.

This dissertation endeavours to produce an informed, coherent and targeted understanding of the feasibility and design of mobile technologies for the self-report of wellbeing and depression during pregnancy. The following chapters describe a complete cycle of research and design defined in part by the triangulation of engagement; informed by comprehensive literature review, alive to the potential of a variety of strategies for the design of engaging experiences, allowing for multiple forms and trajectories of engagement including prolonged and intermittent use of ‘the mindful kind,’ and sensitive to women’s and healthcare professionals’ motivations for interaction with technology. These insights in turn shape a protocol designed to invite and explore the practice of self-report in daily life, facilitating extensive mixed-methods analysis of women’s longitudinal engagement with technology.

This thesis embraces both rigorous quantitative assessment\(^1\) and rich qualitative description\(^2\) of engagement — striving to construct a credible and articulate case for feasibility while also affording the quality of insight necessary to support the action of design. This work advances the adoption of engagement as a nuanced, insightful and challenging descriptor of use with respect to real-world, large-scale, longitudinal and voluntary human-computer interaction.
Chapter 6

Wellbeing in Pregnancy
Women & Midwives

The things we design are not the things but the time spent with them.¹ Mobile devices have the potential to facilitate the self-report and remote screening of wellbeing and depression throughout the antenatal period, extending care to under-served and at-risk populations, enabling timely assessment and intervention, gathering ecologically valid and longitudinal data, overcoming stigma, supporting honest disclosure, and fostering trust between women and midwives. If women and health professionals are to engage with these technologies however, their design must respect women's and health professionals' needs, values, fears and motivations. To date, little design research has been conducted in the perinatal context.

We engaged women, practice and research midwives, medical researchers, clinical studies officers, psychologists, psychiatrists, general practitioners, and clinicians in maternal and child health, obstetrics and midwifery in design with the aim of developing and establishing the feasibility of a mobile application for the self-report of wellbeing in pregnancy. This chapter highlights the experiences of these women and professionals; our context for design.

6.1 Design Research Methodology

Between April 2016 and August 2017, 38 women and healthcare professionals took part in 22 design sessions in the London and Cambridge area.² These sessions were conducted by an inter-disciplinary and mixed-gender team of HCI and public health researchers.³

In partnership with colleagues at Imperial College London, we developed relationships with mid-
wifery clinics and research institutions across the UK through email and in-person over the course of several months. We also worked to recruit women with diverse experiences of pregnancy and well-being, who were currently pregnant or had given birth in the UK within the last 4 years. Women were recruited through a dedicated web-page, the social media accounts of the primary researchers and the Imperial College London School of Public Health, the distribution of cards, posters and emails to medical, community and children's centres across London (See Fig 6.1), through distant acquaintances, and through contact with midwifery clinics and professional organisations across England. Women's partners were invited to attend sessions and an offer was made to reimburse participants' travel expenses up to £15.

Figure 6.1: Recruitment Materials

Fifteen practice and research midwives, 1 clinical studies officer, 1 psychologist and 4 medical researchers/clinicians in maternal health, obstetrics and midwifery each participated in 1 of 5 large design sessions. Seventeen individual sessions were also conducted with 8 pregnant women, 3 mothers, 2 general practitioners (GPs), 1 clinical psychologist, 1 child and adolescent psychiatrist and 2 maternal and child health researchers/clinicians. These smaller sessions enabled the inclusion of women who were unable to travel or who preferred not to discuss their personal experiences in a group setting. Six were conducted over Skype at women's requests. These sessions reflected the reality of participants' daily lives. They were held in university meeting rooms, in women's homes, during breaks at the workplaces of retail managers and stock-brokers, and in the kitchens and quiet spaces of midwifery clinics.

The health professionals who participated in these sessions ranged in age from 25 to 60, had a variety of ethnic backgrounds, and experience working with pregnant women through practice and research. Table 6.1 illustrates the demographic characteristics of our maternal participants. All women
Abbrv. & Week of Pregnancy & Previous Pregnancies & Children & Age (years) & Ethnicity & Nationality \\ P1 & 27 & 0 & 0 & 28 to 32 & Asian British (Indian) & English \\ P2 & Not Pregnant & 2 & 1 & 38 or older & Mixed Other & Singaporean \\ P3 & 22 & 0 & 0 & 28 to 32 & White British & English \\ P4 & 16 & 0 & 0 & 28 to 32 & White Other & French \\ P5 & 11 & 1 & 0 & 38 or older & White Other & American \\ P6 & 35 & 1 & 0 & 33 to 37 & White British & English \\ P7 & 10 & 0 & 0 & 38 or older & White Other & Greek \\ P8 & 31 & 3 & 2 & 33 to 37 & White British & English \\ P9 & Not Pregnant & 5 & 2 & 33 to 37 & White Other & Lebanese \\ P10 & Not Pregnant & 4 & 3 & 33 to 37 & Mixed Other & Singaporean \\ P11 & 39 & 2 & 1 & 33 to 37 & White Irish & English \\ |

**Table 6.1: Maternal Participants**

had received NHS antenatal care. Seven had experienced at least one miscarriage. Two women (P7 and P11) had previously been diagnosed with depression, and one (P5) with anxiety. None had received specific diagnoses of perinatal depression, all professed ‘good’ to ‘excellent’ abilities with technology, were in stable relationships, and held a university or college degree.

Sessions lasted between 1 to 2 hours and were conducted in two parts. The first half of each session sought insight into women's and health professionals’ experiences of pregnancy and perinatal care, entailing group discussions revolving around four themes; experiences of pregnancy and perinatal care, motivations for the self-report of wellbeing, perceptions of self-report and wellbeing technologies, and implementation concerns and preferences (See Appendix A | Design Session Structure.). This process enabled the identification of requirements for a mobile application and study protocol for the self-report of wellbeing and depression. The second half of each session was devoted to concept development and prototype evaluation, engaging participants in the discussion, modification and evaluation of sketches and prototypes at each stage of the design process. Nineteen hours of audio were recorded, transcribed in full, and subjected to thematic analysis. Analysis was conducted both inductively and deductively with respect to the design challenges which arose during sessions, pertaining to wellbeing, perinatal care, self-report, technology adoption and engagement, among others. Participants’ notes were collected and a small number of photos taken.

This chapter highlights the experiences and insights voiced by women and health professionals which informed our understanding of the context of perinatal care and shaped the design of a study protocol for the deployment of a mobile application in antenatal mental health screening.
6.2 Women’s Voices | Experiences of Pregnancy

Figure 6.2: Notes Made by Design Session Participants

The experience of wellbeing and practice of care during pregnancy was our context for design.

6.2.1 Pregnancy | A Bit of a Journey

Women described pregnancy as “a bit of a journey” (P3) punctuated by both positive and negative experiences, "good days and bad days" (P3).

“As a pregnant woman you think in weeks, you really do.” P3

Pregnant women described thinking about pregnancy in terms of multiple concurrent timescales; trimesters, months, and weeks. This patterning shapes parents’ reflections, and provides context for descriptions of key moments; “Oh I felt my first movement at 17 weeks it was so exciting …it was amazing at 14 weeks when I stopped feeling sick” (P3).

This was described by a participant in the 16th week of her first pregnancy as a period of transition and expectation during which “the notion of time is essential but it’s always looking forwards” (P4). Yet, comments from other participants suggest that parents’ reflection comprises past, present and future focus; “Is that the kind of parent I want to be later, and if not, why start now?” (P5).

Pregnancy is a time of change, frequent emotional change, but also change of a more enduring form, to self-identity, and a “shift within the couple’s lifestage” (P4). Emotional changes were described as “hormonal roller-coasters” (P8), crying “and the next minute” laughing (P4), which could lead mothers to feel less ‘emotionally resilient’ (P8). This vulnerability could be compounded by a ‘loss of identity,’ “not at all kind of taken up or kind of considered in your normal healthcare” (P6). Pregnancy, particularly in the first instance, is a significant chapter in parents’ lives, and yet, as one first-time expectant mother expressed, “I don’t want it to be the only thing going on for me” (P1).
6.2.2 Positive & Negative Experiences

When describing pregnancy, women recounted positive experiences, such as ultrasound scans, “really nice shared experiences” (P5) but also did not refrain from expressing feelings of pervasive worry, “when you're pregnant you do worry about everything” (P3). Health professionals expressed an awareness that “everybody is worried during pregnancy, everybody's anxious” (Female GP).

Certain participants described particularly intense and distressing events. A Lebanese mother who had experienced multiple miscarriages described these events as “shocking” and “very devastating” (P9). Her first miscarriage occurred abroad, where she felt the acute absence of a midwife or “any kind of support from family and friends.” Following the subsequent birth of her son, she stated, “I know I had postnatal depression in the sense that I was always scared that I might die and who might look after him, or he might die and how would I feel?” (P9). A second miscarriage, experienced at a later date in the UK, was recounted in similar terms;

“I had no support from the midwife or anyone, because at that point I hadn’t even reached 12 weeks. And I had to do the D&C [dilation and curettage]. I was alone in the UK when I did it. I had my then 4-year old son, he wasn't even 4 years old. So it was very bad [emphasis].” P9

6.2.3 Women’s Concerns

The causes of worry, and less frequently, sadness described by women were diverse. Women expressed concerns related to heightened awareness of their own physical health, “you are so desperate to get [any problem] sorted” (P3), while avoiding “unnecessary interventions, and extra things done” (P5). Healthcare professionals added that “women often say that they don’t feel that they’re doing a good job” and are “very worried that something they’ve done might be a problem” (Female Midwife). These concerns can pervade everyday life;

“You have to bend, and you don't have to bend, and you have to sleep on the right and on the left. All these are worrying because, you know, even when I sleep, I don’t know if I’m harming my baby.” P7

Women named past experiences, “trying for a long time and miscarriages” (P3), as well as knowledge of family history, “my brother died …and then my mum had like a straight of late miscarriages” (P3), as causes of anxiety. Women also expressed concerns for the future, a fear of giving birth, “I've been really struck by others, just quite how terrified they are” (P6), and “scary stories of women not
sleeping, of women not having time even to wash their hair” (P7) in the postnatal period. Worry then
could come to form a vicious cycle, following the belief that “if you worry a lot you could actually cause
yourself to have problems with the baby” (P3).

Co-morbid conditions contributed to many participants’ concerns. P1 recounted ‘a lot of anxieties’
concerning how the fact that she could no longer take evening primrose to help with her polycystic
ovaries would play out. P5 described the attention she paid to avoid ‘burnout’ related to her type 1
diabetes, “how hard it is to do something every day, you know, day in day out,” as well as concerns
related to her high-risk designation due to her age. P9 articulated in great detail the anxiety she felt
when a nuchal translucency screening revealed a risk of Down Syndrome during her 5th pregnancy.

Not all sources of worry and sadness recounted were related to pregnancy. P4 was going through
a period of grieving and was concerned with the impact this might have; “how much the fetus can
feel” (P4). Many midwives pointed to the presence of other children in the home as a common source
of stress. Finally, women spoke of their relationships with their partners. For one woman (P5), the
concern was to avoid a disconnect with her husband by keeping him involved in the pregnancy. For
another (P4), her husband was eager to participate yet she found that “it’s really difficult to find the
place of your partner …for example, a very silly thing, but when you go to midwife there is [sic] no
male toilets. It’s these kind of things.”

These concerns led, for many women, to an urgent ‘need to know’ whether what they were experi-
encing was ‘normal’ — “I wanted to know week on week” (P1), “I just needed to know” (P9).

6.2.4 A Pervasive Stigma

Women and healthcare professionals universally described a ‘massive stigma’ surrounding mental health
during pregnancy. This was articulated in terms of shame (P9), fear (P8, P9), pressure, obligation, and
guilt (P6); “to report that you’re not happy during pregnancy, you’re just not meant to” (P6), “no-one
is afraid to tell their midwife they have pain when they’re pregnant but you are afraid to say ‘well I’ve
been having these dreams …’” (P11).

This taboo ranged from the need to maintain “a certain image of it …when you sit in clinic” (P5)
to a fear of inadequacy as a mother (P4). During their first contact with a midwife, women were of-
ten asked whether their mother had a history of postnatal depression. P4 voiced doubts whether “our
mum’s generation would be able to put that label on it,” and if so how many women “would say to their
daughter ‘Yes when you were born I was depressed.’” These comments mirror previous qualitative re-
search findings which indicate that women often refrain from initiating discussions about their mental
health, or provide inaccurate responses to screening questionnaires, due to discomfort, stigma and uncertainty with respect to the ‘normal’ emotional experience of pregnancy (Byatt et al., 2013; Kingston et al., 2015).

A thread was revealed in women’s comments of a society “very quiet” (P6) on many of the less pleasant experiences of pregnancy, “it’s not what you see on the TV” (P6), resulting in false expectations, and a lack of support when problems do emerge, “I was amazed at how many people said ‘that happened to me too’ but I’d never heard people say it before” (P6), leaving women with the sense of a need to cope alone, “it’s probably hormonal, it’s probably postnatal depression, the fact that I’m always worried” (P9), and subsequently upsetting women’s trust in others;

“Just being told to go and get on with it, and not to worry too much because it was completely normal, but the message that it’s completely normal when you’ve never heard about it or seen, kind of, any references to it, is not very convincing.” P6

Stigma was linked to a fear of particular consequences, most prominently the concern that “if I told them too much they might take my baby away” (P11). Among midwives, this source of apprehension was frequently discussed;

“They talk about the mother and baby homes in the 1980s, that shows you my age …But for them it’s a real threat that you’ll take the baby away from them and I mean I can appreciate where they’re coming from because you get it in the papers ‘oh they’re taking a baby away’.” Female Midwife

Women also voiced distress at the thought of being placed on medication which might harm their child during the antenatal period, “one of the biggest fears I had was that if I told the midwife about those feelings there was a chance they might … put me on a medication that might harm the fetus” (P11), or affect their ability to breastfeed, “if you’ve gotta take whatever drugs the doctor’s gonna give you, that’s gonna go in the breastmilk, and that’s gonna affect your baby’s brain” (P8).

6.2.5 Times of Pronounced Concern

Expressions of concern and distress point to gaps in care and suggest opportunities for screening and support. Women spontaneously described as times of particular anxiety; the planning period (P1), first discovering the pregnancy (P3, P5), the weeks spent waiting to meet the midwife (P3, P5, P6, P7, P9, P11), telling others of the pregnancy (P6), the 30th week (P3), and specific weeks postnatally (P8) (See Fig 6.3).
In particular, the period of time between notifying a GP of the pregnancy and the first midwife appointment was described as “really difficult” (P3), “a time of high anxiety” (Female GP) and “very isolating” (P11). These feelings were exacerbated by a lack of support (P6, P7), morning sickness (P11), bodily changes (P6), pressure to hide the pregnancy (P6), a fear of complications due to co-morbid conditions (P7), and anxiety-induced online searching for information (P3). Miscarriage often occurs during this period, and women described this as “very difficult from a mental health point of view actually, you are just sort of told ‘Sorry, that was a miscarriage, off you go’” (P6). Recognition of the first trimester as a medical and informational gap in line with changes to women's needs over time mirrors previous design research findings (Kraschnewski et al., 2014; Prabhakar et al., 2017).

6.2.6 Experience in Pregnancy

The difference that experience of pregnancy made to women's outlook was striking. Mothers with multiple children on the whole spoke of pregnancy in less emotive terms. Comments made by these women reflect what they had learned and come to expect, “I know, having had two babies myself and expecting my third, that the reality is very much different” (P8).

Experience grants these women knowledge which helps them cope with difficult circumstances, knowledge concerning “how normal those rollercoasters are and how those terrible crazy dreams are” (P11). This is not only knowledge of the emotional experience of pregnancy but of the healthcare system, its pathways, and the likelihood of undesired consequences. These women's experience is a potential resource for first-time mothers but does not always take that form. P5, an American woman in her 11th week, found that maternity advice could be condescending, “you can do pregnancy wrong or you can do it right” and that “as a first time pregnant lady, everyone's like 'oh but don't you know?'.” P6 described feeling “quite rapidly that you're less competent than you were before.”
6.3 The Character & Experience of Perinatal Care

The majority of NHS antenatal care is performed by midwives (See Fig 6.4). A woman’s first meeting with her midwife (the booking appointment) normally takes place during the first 12 weeks of pregnancy. This appointment is typically scheduled following an initial visit to a primary care provider, such as a GP, and can last up to two hours. Eight to twelve appointments usually follow over the course of pregnancy. A small number of appointments are scheduled following birth, at which point a health visitor takes over care. The number of appointments and scans provided varies depending upon local resources and the risk to the mother and child. In certain settings there is a discontinuity of care. Some women will see a different midwife at each appointment. P6 commented that as a result the very notion of ‘your midwife’ proved confusing early in pregnancy.

6.3.1 Midwives’ Roles

Midwives described their goals as ensuring the wellbeing of the mother and baby, monitoring the mother’s vital signs, examining the mother’s past medical history (including mental health), performing a risk assessment, and if necessary making referrals, coordinating care and covering social needs, creating a care plan, reassuring women and their families, providing information, signposting available choices and resources, establishing trust and fostering relationships with women. Midwives therefore have a broad remit and are concerned with both women’s physical and psychological care. Midwives, like other medical professionals, are required to perform composure work, to “remain calm and project an image of efficiency and trustworthiness” while “working on their own threatened composure and over-involvement with the patients” (Strauss, Fagerhaugh, Suczek, & Wiener, 1985). Comfort care in the medical domain is typically “subordinate to survival tasks” but “visible in and around” medical work.
6.3.2 Mental Health Screening

Midwives do not see themselves as mental health professionals; as reflected during discussions concerning the use of the EPDS in-clinic for example; “we’re not qualified to diagnose so it is then up to us to pinpoint the appropriate professional to direct them to” (Female Midwife). Midwives are intensely aware of the treatment gap which exists with respect to PND however;

“Everyone that comes through the labour ward, it’s depression, depression, depression …especially the ones that have lost their babies and …I didn’t have time to speak to one of them …two days later she committed suicide …So I know there is a gap, definitely for people who have lost their babies, there is definitely a gap.” Female Midwife

Perinatal mental health screening is most often conducted during the booking appointment or in the weeks prior to and following birth. The National Institute for Health and Care Excellence (NICE) guidelines recommend screening for anxiety using the Generalized Anxiety Disorder scale (GAD-2) during the early stages of pregnancy, followed by further assessment using the GAD-7 scale, Edinburgh Postnatal Depression Scale (EPDS) or the Patient Health Questionnaire (PHQ-9) if a risk is identified (NICE, 2014). Women recounted significant differences among midwives with respect to their approach to mental health, however.

Appraising Wellbeing as Appearance

Recent research reveals that although almost all UK-based midwives (96%) report asking women about their mental health during their first appointment, only 1 in 10 women recall being asked (Boots Family Trust Alliance, 2013). Almost half of women report never being told about the possibility of mental health problems (ibid.). The women and midwives who took part in this design research presented a similar pattern.

Several women stated that their mental health or wellbeing was rarely discussed, “I don’t know if I had a special case where the midwife generally doesn’t ask anything about your mental health or stress or worries or concerns” (P9), “I don’t think there’s been anything, anything that’s been put in front of me during my pregnancy that has focused on mental health” (P6). One female GP described how “sometimes people don’t want to pick up problems that might take more time” and employ “leading comments like ‘Well you look like you’re fine don’t you? …Things are going really well …you look

(ibid.). P9’s midwife, for example, supported her by writing letters to governmental institutions in an attempt to obtain permission for her parents to travel to England for the birth.
great!’” This line of thought echoes comments made by a community midwife during previous research; “[t]he services are not there to support women and why open a can of worms that you can do nothing about” (Boots Family Trust Alliance, 2013).

Appraising Wellbeing Indirectly

Women also described an indirect, conversational approach employed by some midwives, “they might ask you how you’re feeling but that’s a really general question isn’t it?” (P8). This light-touch approach is most likely motivated by midwives’ appreciation of the stigma surrounding mental health. It is therefore worth noting that women frequently articulated awareness of midwives’ intentions, “I know what they’re asking…” (P10), “that is the question that they were trying to ask” (P9).

Appraising Wellbeing Directly

Another approach recounted by women comprised a more direct form of questioning, including verbal application of the Whooley Questions. Some women found this approach “so weird, I’d never met them before, I had no idea what I was doing being pregnant” (P6). Two women (P2, P3) compared this to a routine question concerning domestic abuse which also took them by surprise. Midwives also expressed doubt whether this approach does “probe as deeply as perhaps is needed sometimes.”

6.3.3 The Midwife-Patient Relationship

“I think midwives care really deeply about getting a high quality baby out and the mommy being ok.” P11

The relationship between a woman and her midwife is a characteristic feature of antenatal care and an essential factor in the success of mental health screening. Women described midwives as “very supportive and nice” (P9), “amazing” (P8), and “nice people, very lovely people” (P10) who “care really deeply” (P11). Women’s perceptions of midwives’ roles coincided for the most part with midwives’ own accounts; a support role, “there more to aid you through” (P3).

Two women expressed the sentiment that “a midwife is not a mental health professional” (P11), “it’s not that I would have wanted her to [ask me about my mental health], I never expected her to, I never thought that was something they actually covered” (P9). Instead midwives were described as generalists, a disappointment for P9, “all the reasons for nuchal translucency …even when I spoke to the midwife, I couldn’t really talk to her about these things.”
Women’s reflections highlight those characteristics of the midwife-patient relationship which most strongly influenced their willingness to disclose concerns;

Routine

Midwife appointments were described by some women as ‘very routine,’ “more concerned about taking measurements,” “you go in, and you do the samples, and you do the tests, and it’s like ‘see you next time’ ” (P2). For P4, a woman in the 16th week of her first pregnancy, this was not a positive experience; “after my first meeting with my midwife I was very upset about this … look into my wellbeing, I’m a mother, and I’m not just a number, I’m not just … a blood test.”

Risk Labelling

Women also felt that their initial classification as low or high-risk tended to dictate all subsequent interaction with their midwife. Women classified as low-risk were critical of this approach; “low-risk, from a medical perspective is not the same as wellbeing, and you still have questions” (P4).

Time Pressure

NHS midwives describe significant time pressure as a constant feature of their work, and many women expressed awareness of this challenge, often in sympathetic terms. However, this pressure also impacted women’s willingness to disclose concerns; “they basically just kick off my appointments with ‘Oh great, you’re entirely normal, this is gonna be really quick’ ” (P6), “I did have questions for her but I just felt like she was in a rush, I wasn’t particularly comfortable in the environment” (P1).

Communication Styles: Unreasonable Positivity

Another factor which women described as detrimental to their relationships with midwives was an overly positive, ‘incredibly cheerful’ (P6) attitude; “everything that came out of that midwife's mouth was like ‘perfect perfect perfect,’ so I was very reluctant to tell her the truth” (P11). Women’s comments reveal a fine line between reassurance and disregard. P9 did not take positively her midwife's statement that ‘oh baby’s fine, you’re fine’ when she expressed concerns regarding her weight. Whereas P8, an experienced mother in her fourth pregnancy, spoke of similar comments as reassuring, “they don't really care, because, in fact they probably think that's a good sign.”
6.3.4 Technology Use in Pregnancy

Women frequently spoke of the role technology played during pregnancy. Every participant owned a mobile phone, and 6 owned tablet computers. Every woman had used many websites related to pregnancy, and all but one (P7) had used mobile applications, including BabyBuddy, an app comprising a goal and appointment tracker, educational articles and other features, the NHS approved app Pregnancy+ which provides daily updates throughout pregnancy, a personal diary, weight log, kick counter, contraction timer, baby shopping list, baby names and images of prenatal development, Mind The Bump, a mindfulness meditation app for new parents, Squeezy, an NHS approved app for pelvic floor exercises, BabyCenter, WhatToExpect, Emma’s Diary, and NetMums.

Many women used email (P8, P9) and phone (P1, P3, P8) to contact their care providers. One participant used her GP’s phone service alone to inform her of her pregnancy (P3). P9, while travelling abroad, remained in touch with her UK-based gynaecologist by email. A well-timed notification from an app helped P3 identify a water infection. P6 and P1 described using app content as a way to engage their partners in their pregnancy, “a sort of third party thing that you can point at and say ‘Look, this is what might be coming up next’.”

Women spoke most frequently of technology as an important source of information, employed to overcome perceived shortcomings in care. This use of technology for information seeking was described by women during previous research as a means to compensate for “useless” and “overwhelming” printed literature (Peyton, Poole, Reddy, Kraschnewski, & Chuang, 2014b). The use of technology was not always recounted positively however. P6 found herself targeted with online advertising related to her pregnancy, “Facebook notices you’re pregnant … Google also knows I’m pregnant now.” Searching for information online could be overwhelming, “a pit of anxiety,” a “terrifying free-for-all” (P6), and the removal of international barriers did not always help, “you can read things that just make you more anxious because you don’t know why you’re not getting that care” (P6).

6.4 Designing a Feasibility Study

“Studying technology in its social milieu brings about great challenges as to how one can account for its ‘messiness’ and at the same time can be able to accurately answer the research questions.” DePaula, 2003

This insight into women’s and healthcare professionals’ concerns, values and motivations shaped a protocol for the deployment of a mobile application in antenatal mental health screening. This would
be the first use of a mobile application for the longitudinal self-report of wellbeing in daily life in the context of pregnancy and would require integrating mobile technologies within routine NHS pathways. The design of this protocol was informed by what we have learned of the characteristics of multidimensional, salutogenic and pathogenic measures of wellbeing (See Chapter 2), the importance of assessing experience with respect to multiple time-frames (See Chapter 3), the strengths and weaknesses of previous EMA protocol designs (See Chapter 4), the need to triangulate measures of engagement in daily life, over time and with respect to voluntary use (See Chapter 5), the lived experience of women and healthcare professionals (See Chapters 6-7), and our duty of care.

6.4.1 Aims & Constraints

The primary aim of this study would be to assess the feasibility of a mobile application running on women’s own devices for the repeated and longitudinal self-report of psychological wellbeing in pregnancy. It would also examine the role of mobile technology as a means to address barriers to antenatal mental health screening, the collection of momentary, experiential, ecological and retrospective data throughout pregnancy, and critical success factors for the deployment of self-report technologies, including user engagement and adherence, with respect to two distinct sampling protocol designs.

6.4.2 Protocol Design

Retrospective assessments would take the form of the Edinburgh Postnatal Depression Scale (EPDS), a validated screening tool employed to identify women at risk of depression both during and following pregnancy, sensitive to changes in the severity of depression over time, recommended by the NICE guidelines, and familiar to midwives (Cox, Holden, & Sagovsky, 1987). This 10-item self-administered survey instrument screens for the presence of clinical symptoms including guilt, sleep disturbance, reduced energy levels, anhedonia and suicidal ideation over the past 7 days. A woman’s response to each question is recorded on a scale ranging from 0 to 3 points.

Momentary assessments would consist of five questions concerning participants’ mood, rest, worry, enjoyment and energy in the moment, and two contextual questions pertaining to participants’ present location and activity. The choice of EMA constructs was informed by the work of a research fellow in clinical psychology at the Collaboration for Leadership and Applied Health Research and Care for the East of England (Blake, 2014) as well as the extant literature examined in Chapters 2 to 5, and was corroborated by design research participants.5

Participants’ use of the mobile application would be logged throughout the study, and this quanti-
tative data used to support the assessment of feasibility, protocol compliance and data validity, including drop-out rates, adherence to sampling protocols, usage patterns, timeliness of data collection, data completeness, session duration, and the use of app features.

Women’s qualitative impressions would also be gathered two weeks following their final notification, in order to rule out potential confounds, triangulate a subjective measure of engagement, and gauge the experience of use. A questionnaire containing up to 14 questions (depending upon the study arm) was created, drawing on the Usefulness, Satisfaction and Ease of Use (USE) scale and including additional questions concerning women’s desire to continue use and the value of self-report.

6.4.3 A Two-Arm RCT

A randomised controlled trial (RCT) study with two arms was designed to enable comparison of women’s engagement between a retrospective assessment strategy and a retrospective plus momentary assessment strategy. This would also permit analysis of the value of momentary assessments of wellbeing and the relationship between momentary and retrospective reports.

Participants in each arm would be prompted to provide reports over a period of 6 months during pregnancy. In the first condition, participants would be requested to provide retrospective reports once per month through a mobile app. Participants would receive their first notification 2 days after installing the application, a delay chosen to give users time to familiarise themselves with the app without its presence fading from memory. Subsequent notifications would be randomly generated between the hours of 17:00 and 21:00, 21 to 35 days apart, for a total of 6 notifications over a 6 month period (See Fig 6.5). In the second condition participants would be requested to complete both retrospective and momentary assessments. Monthly sampling periods would consist of 6 contiguous days of semi-random assessments, comprising a single EPDS prompt between 17:00 and 21:00 on day 1, followed by four days of momentary assessments 3 times per day (between 09:00–12:00, 13:00–16:00 and 17:00–20:00), and concluding with a final EPDS assessment between the hours of 17:00 and 21:00 on day 6. All partici-
pants would again receive their first notification 2 days following installation of the mobile application. Prompting would be repeated 21 to 35 days following the previous period, for a total of 6 assessment periods. Participants would be free to provide reports, ignore or disable notifications, or delete the app at any time. No reminders or follow-up notifications would be sent to participants, and no monetary incentives would be provided.

6.4.4 Recruitment

<table>
<thead>
<tr>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 years old or older</td>
<td>Diagnosis of any common mental health disorder (i.e., depression or anxiety disorders) as specified in the DSM-IV</td>
</tr>
<tr>
<td>Attending antenatal clinics in participating GP practices</td>
<td>Receiving treatment for any common mental health disorder</td>
</tr>
<tr>
<td>Own an Apple or Android mobile phone</td>
<td>Recent personal history of any common mental health disorder (i.e., within the past 12 months)</td>
</tr>
<tr>
<td>Are in the first trimester of pregnancy</td>
<td>Not comfortable reading and writing in English</td>
</tr>
</tbody>
</table>

Table 6.2: Participant Inclusion & Exclusion Criteria

The participants of this study would be pregnant women attending NHS antenatal clinics in general practices, community services and secondary care centres across England, recruited during the first 14 weeks of pregnancy. This limit was chosen to ensure that the majority of assessments would occur during the antenatal period. Additional inclusion and exclusion criteria (See Table 6.2) were informed by the recommendations of Thombs and Ziegelstein (Thombs & Ziegelstein, 2014).

Recruitment would be conducted by members of the clinical care team with an additional research role (research midwives), or by Clinical Studies Officers (CSO) with an honorary contract or Letter of Access from participating NHS centres. Potential participants would be approached opportunistically in the waiting areas of participating centres and assessed against inclusion and exclusion criteria by a member of the clinical care team through a review of medical notes, by the CSO through discussion with the clinical care team, or by asking potential participants directly.

Interested and admissible women would be provided with a participant information sheet, have the details of the study explained to them, and be presented with the opportunity to ask as many questions as desired. Women would be granted a minimum of 24 hours to decide on participation and refusal to take part would have no impact on their legal rights, medical care or relationships with care providers. However, if participants wished to take part on the day they were approached, they would have the opportunity to provide written informed consent. The recruiter would then ask the participant to complete a personal demographic survey, the Whooley Questions7 and the EPDS using a tablet provided
by the research team. This demographic survey contains 11 questions pertaining to participants’ age group, ethnic background, marital status, employment status, level of education, smartphone and tablet computer ownership, obstetric history, and personal history of depression. The recruiter would then guide the participant through the process of downloading (from either the Apple App Store or Google Play Store) and installing the mobile application onto their own device. When possible, installation would take place within the clinic. However, where internet services were not available, participants would be provided a userID with which to install and register the application in their own time.

We chose to relate the proposed sample size to a 95% confidence level and a confidence interval of 10. We would therefore require 96 participants in each experimental condition, to which women would be randomly allocated; resulting in a total sample size of 192 participants.

6.4.5 Responding to Alerts | Fulfilling Our Duty of Care

All participating women would be prompted to complete the EPDS in daily life. Aggregate scores between 10 and 12 points on this scale suggest an increased risk of depression (Allbaugh, Marcus, Ford, & Flynn, 2015). Scores of 13 points or above indicate that the diagnostic criteria for major depressive disorder have most likely been met, and question ten queries self-harm and suicidal ideation. This study design therefore entailed considerable ethical and medical responsibilities, and a significant duty of care towards participants.

The central research team would agree on a list of designated contacts with each clinical care team prior to the study, including clinicians on duty during and outside of normal working hours (including weekends and bank holidays). We would also collect contact details for every participant’s GP and designated clinician at baseline. Throughout the study, women’s encrypted and pseudo-anonymised data would be automatically synchronised with a research server over a secure connection. A colour-coded pseudo-anonymised alert would be communicated to the research team by email should any participant’s data meet the following criteria;

- Yellow (Mild) Completed more than two EPDS assessment on the same day, with an EPDS score of 9 points or less and a score of 0 on question 10.
- Orange (Moderate) An EPDS score between 10 and 12 points, with a score of 0 on question 10.
- Red (Severe) An EPDS score of 13 points or more, or 1 or more points on question 10 regardless of the overall score.

In the case of a red or orange alert, the study coordinator would contact the designated member of
the participant’s clinical care team by phone and email within 24 hours. The clinical care team would then follow up directly with the participant.

6.4.6 Ethics

This study protocol was reviewed and approved by the National Research Ethics Service Committee South East Coast-Surrey on the 15th of April 2016 as a notice of substantial amendment to an original submission (9th July 2015) under the Research Ethics Committee (REC) reference 15/LO/0977. This study was sponsored by Imperial College London under the reference number 15IC2687 and included in the UK Clinical Research Network Study Portfolio under the Central Portfolio Management System number 19280.

6.5 Conclusion | Reflections & Takeaways

This chapter presents insight into the experiences of women and healthcare professionals during pregnancy and describes the design of a protocol to support the first clinical deployment of a mobile application for antenatal mental health screening. This study design entails a longitudinal time-frame, contrasting notification schedules, momentary and retrospective measures of wellbeing and depression, a diverse, voluntary and nationally-distributed population group, an extensive mixed-methods approach to the assessment of engagement, and the urgent provision of care following reports of depression, self-harm and suicide risks.

The feasibility of any technology ultimately hinges upon its design. Chapter 7 describes how these sessions further informed the design of a mobile application for the self-report of psychological wellbeing and depression during pregnancy.
Chapter 7

Designing Self-Report in Pregnancy | BrightSelf

Good design makes good science possible.¹ This chapter describes the design of a mobile application for the self-report of wellbeing and depression during pregnancy; BrightSelf.

7.1 The Design Brief

“Although psychologists have developed many proven ways to strengthen our mental resources, we spend much more time with digital technologies than we do with psychologists.” Calvo & Peters, 2014

Mobile devices have long been thought to possess the potential to transform research, clinical practice and wellbeing at a population scale (Shiffman et al., 2008; Lazer et al., 2009; Dufau et al., 2011; Atkins & Frazier, 2011; Wichers et al., 2011). Millions of women have installed thousands of mobile applications in the hope of supporting a healthy start to life. The majority of these technologies have been designed to communicate health-related information to parents (Tripp et al., 2014). HCI researchers have developed prototype applications for Dutch (Babywijzer), Pakistani (Baby+) and Vietnamese Australian (We-HELP) populations (Wierckx, Shahid, & Al Mahmud, 2014; Sajjad & Shahid, 2016; W. Smith, Wadley, et al., 2017), deployed SMS-based systems for personalised health information communication in Kenya and Pakistan (Perrier et al., 2015; Batool, Razaq, Javaid, Fatima, & Toyama, 2017), and conducted qualitative analyses of pregnant women’s motivations for information sharing and support seeking online (Kraschnewski et al., 2014; Peyton et al., 2014b; Gui, Chen, Kou, Pine, & Chen, 2018). Peyton et al. propose a ‘pregnancy ecology,’ comprising physical, emotional, informa-
tional and social supports, to support the design of physical health interventions (Peyton et al., 2014a), which Prabhakar et al. extend to include support needs, sources and interventions within an Evolving Ecology of Support (Prabhakar et al., 2017).

Researchers have also explored the design of prototype technologies for health data tracking in pregnancy, including nutrition, hydration, activity, weight and mood (Bloom) (Wenger, Bell, McEvoy, Yamaguchi, & Shokrpour, 2014), nausea and vomiting (Dot-it) (T.-I. Lee, Chiang, Guo, Chen, & Chen, 2016), and physiological data (Nuwa) (Gao, Li, Lin, Liu, & Pang, 2014). Other perinatal research has examined women’s motivations with respect to menstrual tracking applications (Epstein et al., 2017) and the disclosure of pregnancy loss on social networks (Andalibi & Forte, 2018), the prediction of postnatal depression from survey data (Natarajan et al., 2017), the use of twitter to track development milestones in young children (Suh, Porter, Hiniker, & Kientz, 2014), and the design of applications for monitoring the health of preterm babies (Estrellita) (G. R. Hayes et al., 2014), sharing infant activity data with friends and family (MammiBelli) (Hui, Ly, & Neustaedter, 2012), and to support breast-milk donation (Milk Matters) (Wardle, Green, Mburu, & Densmore, 2018).

Design research has facilitated knowledge of women's needs with respect to a variety of prototype systems. However, less HCI research has attended to the real-world use of technology in the perinatal context, the role midwives and other health professionals play in pregnancy, the integration of personal devices within a clinical and public health context, the subjective experience of pregnancy and the significant implications of maternal mental health, nor has progressed beyond low-fidelity paper prototypes.

Our aim therefore was to develop a mobile application for the self-report of wellbeing and depression during pregnancy. This technology would be required to engage women in self-report over a period of 6 months or more, support the collection of valid retrospective and momentary reports of mood, rest, worry, enjoyment, energy and depression in daily life, as well as manage depression, self-harm and suicide risks. This system would be deployed within a public health service for longitudinal and voluntary use by a highly diverse demographic group. This would entail design to facilitate disclosure, shape the self-report of subjective experience, make sense of wellbeing, provide value to women and midwives, avoid medicalising normal emotions and refrain from undermining the human component of care.

7.1.1 BrightSelf | The System Architecture

The resulting system, BrightSelf, comprises mobile applications for Android and iOS, a server for data storage, management and alert-provision, and an online interface (See Fig 7.1).
BrightSelf provides a platform for the self-report of wellbeing during pregnancy; retrospective reports in the form of the EPDS (a validated screening scale for perinatal depression) and momentary reports in the form of visual analogue scales for mood, rest, worry, enjoyment and energy, as well as two questions concerning location and activity context. Each app comprises two versions, one of which is activated according to the userID entered upon registration. The first supports retrospective reporting alone (Arm 1), and the second both retrospective and momentary reporting (Arm 2). Additional features include interactive visualisations of users’ data, information regarding perinatal wellbeing and the study itself, contacts for mental health support, and an ‘ideas machine’ designed to support longitudinal user engagement. Accessing the app requires a 4-digit PIN chosen by the user during registration. The app logs the time users spend on every screen, the use of every significant feature, the device type, and each user’s uniquely generated notification protocol.

This study design entailed significant ethical and medical responsibilities given the collection of data reflecting women’s risk of depression, self-harm and suicide. Data stored on women’s devices is encrypted using 256-bit AES encryption and synchronised with the study server using a Secure Socket Layer (SSL) connection throughout the study. A colour-coded alert is instantly communicated to the research team by email when any participant’s data meets EPDS risk criteria. A separate offline database containing identifiable clinician and patient data is located within the School of Public Health at Imperial College London, encrypted, and accessed through a policy of authorised usernames and passwords.

7.2 Designing BrightSelf | A Mobile App for the Self-Report of Psychological Wellbeing in Pregnancy

“The design process is composed of a series of decisions, which pare down the vast set of potential systems to the one that is actually delivered.” Dix, Finlay, Abowd, & Russell, 2004
The design of BrightSelf drew on the lived experience of women, midwives, GPs, clinical psychologists and other health professionals, the public health, wellbeing, cognitive psychology and HCI literatures, and the experience of the research and design team. This was an iterative process which included the evaluation of sketching and prototype applications during design sessions with users.

This section describes many of the choices entailed in the design of BrightSelf, and their implications for future health and wellbeing technologies. I encourage the reader to visit the Appendices for detailed images, screen flows and additional design rationales (See Appendix B | Design Work).

7.2.1 The Importance of Appropriate Framing

“How do people position it? How do you want to position yourself?” P4

These design sessions highlighted women’s intense need to understand their wellbeing during pregnancy, and the challenges faced by the health professionals attempting to do so. Tensions of personal information and public health necessitate the careful framing of any technology introduced into a clinical context.5

An Aversion to Mental Health

“A lot of people, as soon as you say the word mental health, shut up, completely.” P4

Women described themselves as low, anxious and in need of support, and yet were keen to dis-
tance themselves from any labels related to mental health or illness. An intense disconnect with the term 'mental health' suggests that employing this language might reduce the reach of health and well-being technologies (Matthews & Doherty, 2011). While some women preferred the term 'psychological wellbeing' others felt it possessed similar negative connotations, "psychologist, psychiatrist, mental ... for me, sit in a similar category" (P8). Several women spontaneously proposed 'emotional wellbeing' as a preferable alternative, and these discussions informed the tone of the language employed within BrightSelf.

The first prototype was assigned the name ‘BabyBright.’ A female researcher and several midwives commented, however, that the inclusion of the term 'baby' in the title drew attention to the link between maternal mental health and the sense (or loss) of autonomy experienced by pregnant women. An iterative design process led to the revised title BrightSelf, chosen to present the app as an intimate resource with a light touch and the potential to facilitate self-knowledge (See Fig 7.3).

Every Pregnancy is Unique

Pregnant women and midwives are both highly heterogeneous groups. Fourteen midwives took part in design sessions 2 to 5, during which discussion often focused on the need to address tensions between the provision of care and pragmatic issues pertaining to the unique character of women's daily
lives, including transport, affordability and even access to mobile applications. Midwives emphasised the need to avoid assumptions concerning the ownership of mobile devices, purchasing power, or expectations of privacy and autonomy in regard to communication.

“...You’re going to get many of the same kind of women ...motivated, articulate, intelligent, interested in their own wellbeing ...and the ones that we really want to target ...probably don’t have ...access to ...and you think ...of our Asian population ...their mobile phones ...actually their husbands have the phones.” Female Midwife

These comments illustrate the value of eliciting feedback from professionals who work closely with under-served communities and are therefore well-placed to provide insight pertinent to design, particularly in the case of limited access to vulnerable users. This approach embodies ethical pluralism, recognising the co-existence of value systems and the capacity of localised phronetic knowledge to support design (Barry & Doherty, 2017). When designing perinatal technologies, it is often essential to avoid assumptions pertaining to pregnancy planning, family history, the presence of a supportive partner, infant development, other children, miscarriage or the absence of complications and co-morbid conditions. Every pregnancy is unique.

Introducing an App

The BrightSelf mobile app is introduced to women during their first appointment with a midwife. This provided an opportunity to scaffold the framing of the application. Both women and health professionals voiced the need for this to be a normalising experience; “same as you go to see [the] midwife for the booking appointment, it's just something that everyone does” (Female Psychologist).

We designed an illustrated guide to recruitment, a 20-page A5 booklet for each study arm, containing concise description of the system in large text, full page screenshots, and streamlined instructions for recruitment, installation and use (See Fig 7.4). This shared object allowed midwives to walk women through the use of the app and the aims of the study in an efficient, straightforward and standardised fashion. BrightSelf was described with a light and positive tone and framed as empowering the individual patient.

This interaction illustrates how the pathways of clinical care constrain design in the perinatal context. Although women highlighted the weeks prior to their first appointment with a midwife as a time of particularly intense anxiety, there is currently no opportunity for midwives to interact with patients during this period, and therefore, even were women provided a screening tool, no pathway exists to support subsequent action.
7.2.2 Supporting Value in Self-Report

“I’m not going to just use it because you ask me to use it, right.” P10

Women often made comments reflecting a need to perceive value in the practice of self-report before they would accept the potential burden of reporting; “What’s in there for me?” (P4). Understanding women’s motivations for engaging in self-report was therefore essential to the design of BrightSelf. Women expressed motivations pertaining to their own subjective experience, relationships with others, and the potential for action.

Turning In | Understanding Self

Many women perceived value in the capacity of self-report to enable and maintain self-knowledge, progress and wellbeing-related change; “you notice a pattern and you know yourself better” (P10), “if you’re tracking your own mood, it tends to make your own mood better” (P11), “if you keep a diary I think you can see if that is working for you” (P7). Less anticipated were those motivations expressed by women which related simply to supporting self-awareness and reflection; reminiscing, comparing trimesters, checking in “treating it like a game” (P5), venting “to someone or something” (P9), acknowledging emotion and preserving experience, “in the middle of the night when you wake up …if I could
write it down, sometimes that’s out of your head” (P1), “for people to be able to acknowledge how afraid they are” (P6), “it might be quite useful in your second pregnancy …to have some of that information” (P3). One participant (P11) proposed devising a ‘niggle chart,’ a “map of a cartoon person,” which a woman might then annotate with her physical and emotional experience. This range of motivations reflects the potential variety of meaning and action associated with the phrase ‘technology-mediated reflection’ (Mamykina et al., 2008; Isaacs et al., 2013).

Turning Out | Understanding Self & Other

Women’s motivations also reflected the socially situated reality of pregnancy. Women envisioned the use of data to both avoid feeling alone, “the reason why everyone Googles so much” (P3), and to feel normal, “Is this normal? Am I normal?” (P10). Healthcare professionals often spoke of women’s need to know what’s normal, an important factor in the success and failure of mental health screening programs; “they know they’re struggling but they’re not sure if everybody …if that’s normal” (Female Psychologist), “we get asked that, is this usual?, are these feelings usual?” (Male Maternal Health Clinician). One participant (P8) even suggested that a woman’s birth partner might be better placed to evaluate her wellbeing, “you could say ‘this is an app for you, to help look after me,’ it’s almost like ‘mother the mother.’” In a recent UK-wide survey, 90% of perinatal health professionals reported a belief that women’s partners would spot signs of deteriorating health before she herself might, although fewer than 1 in 5 reported meeting partners as a key part of their work (Boots Family Trust Alliance, 2013).

Taking Action

Women also considered the potential for data to support action as an incentive for its collection; inciting conversation, “to ask the right questions” (P4), inspiring action, “I’m suffering from something, I should do something about it” (P10), serving as a means to obtain support, and enabling midwives to tailor care, “to really move away from high and low-risk pregnancy” (P4).

Reservations

Finally, several participants also expressed concern that engaging in the practice of self-report might reinforce negativity, “writing something down is a very firm commitment” (P5), “creating a diary of depression” (Female GP), lead to obsession, “I’m not really one of those ‘FitBit’ people” (P8), or bias reflection, “only when you’re desperate and feeling really lousy then you go on it” (P10).
7.2.3 Shaping a Better Self-Report Experience

What we report of our wellbeing is invariably shaped by the experience of self-report. The methodological concern of reactivity, for example,\(^8\) translates into a design constraint for self-report technologies. The collection of valid data also hinges upon the engagement of users over time. We sought to address this critical design challenge, in part, by designing an efficient and enjoyable self-report experience, employing text, animation, illustration and unique, responsive forms of interaction (See Fig 7.5).

Participants expressed strikingly different responses to the experiences of retrospective (EPDS) and momentary (EMA) reporting during prototype evaluations. The medical tone of the EPDS often swiftly evoked guarded reactions among women, “see, this is all the mental health stuff…” (P5), “now I’m being like ‘Ooo I’m being evaluated, so I need to be careful’” (P4). Question four of the EPDS, which asks women whether they have felt ‘anxious or worried for no good reason,’ often led participants to exclaim “what if you’re anxious for a good reason” (P5), “women hate it!” (Female Psychologist). Questions concerning ‘sad or miserable’ feelings and crying were also described as difficult to complete during pregnancy, and associated with negative, stigmatised and non-negotiable consequences.

In contrast, the EMA reporting experience was described as “really easy and not intrusive” (P6), ‘light-hearted’ (P3), “really nice and almost a bit fun as well” (P1). These momentary measures support reflection with respect to a broader spectrum of emotional experience, including positive dimensions,
in comparison to the EPDS. This may allow women to present a more complete appraisal of their well-being, and to view reporting less as “a test that you pass or fail …a box tick exercise,” as one female GP described the EPDS. Effective mental health screening programs are required to balance the sensitive presentation of avenues for disclosure with the need to survey populations and pro-actively identify those at risk who might not otherwise feel able or sufficiently self-aware to seek support. The content and presentation of the EPDS was necessarily constrained in comparison to the design of the visual analogue scales employed for momentary reporting. Women reported that they would complete EMA reports more often and more honestly, underlining the importance of user-experience design for the self-report of wellbeing.

Terms such as ‘EMA,’ ‘momentary,’ ‘retrospective’ and ‘assessment’ reflect medical and academic framings less likely to support honest disclosure. Discussions with women and health professionals revealed a preference for the terms ‘Check In’ and ‘Check Back,’ which better express women’s autonomy while highlighting the orientation of each form of self-report with respect to time (See Fig. 7.6). This is design to evoke distinct momentary and retrospective selves (See Section 3.9), accomplished in part by distinguishing modes of self-reflection by means of unique and consistent colour palettes, iconographic schemes, concise and carefully chosen text, interaction mechanisms, styles of data visualisation, and exclusive forms of automated feedback.
Design researchers have long stressed the need for health-tracking technologies to provide feedback to users (Kraschnewski et al., 2014). We were similarly keen to provide feedback to women immediately following each experience of self-report; an opportunity to provide context for each woman’s data, support self-insight, and incentivise repeated use. Midwives, however, cautioned against labelling women’s subjective experience, stemming from a desire to avoid misrepresenting their own role as that of a mental health professional and to avoid simplistic, and possibly erroneous, remote interpretations of subjective experience; “they might think ’well hold on a minute, I don’t feel that’” (Female Midwife). As Calvo & Peters write, “’be angry’ is poor direction for a stage actor, and ’don’t be so sad, cheer up!’ is fairly useless advice from a well-meaning friend. Technology designers should be wary to avoid the similar trap of trying to tell people what to feel” (Calvo & Peters, 2014).

In particular, there was a reluctance among some midwives to present the score generated by the EPDS to women, a feature of early prototypes. Midwives voiced concern that this could lead women to “try to beat their score” or “feel pressured to be in that range.” Conversations with women, often in contrast, reflected a need to facilitate meaningful interpretation of their data and to accord value to the practice of self-report. Compromise was eventually realised, following an extensive process of iterative prototyping, in the form of non-numerical representations of EPDS scores, simple animated visualisations of women’s data, and automated semi-random feedback endorsing self-awareness, self-
care and support-seeking behaviours according to each woman's state of wellbeing (See Fig 7.7).

7.2.4 Facilitating Self-Awareness & Disclosure

In a recent UK-wide survey, less than a fifth of women reported being completely honest with their midwives (Boots Family Trust Alliance, 2013). Women described a reluctance to talk about their feelings as the most significant barrier to care (ibid.). This data suggests that seeking mental healthcare during pregnancy requires women to recognise their symptoms, acknowledge their need for support, and trust the services and individuals they might approach for help.

A Safe Space

Several participants made comments suggesting that technology might make it easier for women to disclose how they are feeling; “through an app I might be a bit more open” (P1), “a safe space where you can articulate some of the anxieties that you’re feeling” (P6).

Creating a safe space requires establishing trust between women, midwives and, in this case, researchers. One woman who had previously experienced depression spoke of her hope that sharing data might serve as an opportunity for her midwife to express care and reassurance, articulating her sense that, in the past, nobody had “really show[n] any interest in …how I feel …nobody asked” (P7). She also commented however that her willingness to share how she was feeling would depend on a perception of her midwife as 'sensitive,' “interested in mental health issues,” and able to “assess this, in a more calm way maybe, more rationally.”

Women spoke of significant variance among midwives. One woman explained “you get the 'earth mother’ types, and then you get the more 'clinical matron' types but …they tend to be very much going by the way they do things, and they’re not really that regimented in this sense” (P8). This participant describes midwives not as “data driven professionals, like you might get clinicians or consultants are, or scientists …the midwives I’ve worked with, they are, emm, they’re not really on the ball with this sort of stuff” (P8). Several women were therefore sceptical of midwives' abilities to process their data. P5 spoke of how her diabetes and insulin related data was used “very ineffectively” by her nurse and diabetic obstetrician, stating that “its overwhelming for practitioners” given “the amount of time that they have.” Participants also indicated however that they would not want their midwife to act as a purely data-driven professional who might say “ ‘You scored 10 out of 10, good one!’ I don't want to have this kind of chat with my midwife” (P4). For this woman, sharing information concerning her wellbeing with her midwife would prove useful only if “the midwife has received training, and when I’m talking
about training, I’m talking about therapeutic training, about how to handle with care the data” (P4). Midwives also expressed a need to balance the information sought from women against their capacity to take action based on those disclosures. For the purposes of this feasibility study, women’s EPDS scores alone would be communicated to midwives following reported risks of depression, self-harm or suicide.

Transparency is a key factor in design to support help-seeking — transparency pertaining to the limits of what it is possible to reasonably infer from different forms of data, to how women’s data is interpreted, by midwives in isolation, in conversation with women or by employing particular sense-making strategies for example, and to the potential consequences of disclosure. As one participant states, “I think the reality is you know someone’s not going to take your baby away if you tell them that you’re feeling this way but you don’t know whether it means a midwife’s going to come to your house every other day ... when you’re in a very anxious moment or in a low period you can’t rationalise” (P11). Studies show that health data of all kinds is not ‘self-evident’ but subject to “manipulation and interpretation based on context and experience,” and that “work must be done in order to use the data in practice” (Mentis et al., 2017). Women and midwives often spoke of the importance of interpretation;

“Let’s say ‘I’m upset today because my son upset his room, for a very stupid reason, ok’ and I write ‘I’m upset today,’ and the next day, for example, I have a fight with my husband, and the third day, I have a fight with my friend, and the fourth day, whatever, I got a really high electric bill, ok, so this looks like a really bad week, and then the midwife, when I meet her ... the midwife might say ‘Oh well look, there’s a trend here.’ Maybe there isn’t a trend [emphasis].” P9

One approach to minimising the potential for such misunderstanding is to grant users the opportunity to provide sufficient context to preserve the meaning in their reported data; “I would want to write the word ‘scan’ under that, or ‘hypo’ under the ones where I’m worried, anxious” (P5). The systematic interpretation of open-ended data also presents clinical and ethical challenges however. Women recounted causes of distress which were not always related to pregnancy and which fell outside the remit they were comfortable granting to midwives;

“Bad mood. So then she sees a negative thing and is like you know ‘Why did you?’ ‘Well, you know I had a fight with my husband,’ and maybe that is something you don’t want to share with your midwife. And it has nothing to do with your pregnancy.” P9

Women spoke of conflicting needs for accuracy and privacy related to the potential interpretation, and over-interpretation, of their data. An appropriate level of detail must therefore balance midwives’
responsibilities with the onus on women to explain the reports they make. Although midwives were not required to access women's momentary data as part of this feasibility study, BrightSelf did include two brief and lightweight multi-choice assessments of context following each momentary report; 'where you are' and 'what you're doing,' in order to facilitate analysis of the resulting corpus.

As many EMA studies have found, engaging users in self-report typically requires reducing the burden on users while also providing intrinsic and extrinsic incentives for use (See Chapter 4 & Section 5.6.9). BrightSelf includes a variety of supplementary features designed to engender trust, inspire engagement, and support users. This includes interactive visualisations of users’ data, information regarding perinatal wellbeing, the study and the mobile app, as well as contact details for numerous sources of mental health support (See Fig 7.8).

BrightSelf respects women’s autonomy and ownership of their health and wellbeing by granting complete access to their data and control over the visualisation and interpretation of their mood, rest, worry, enjoyment, energy and EPDS scores, graphed either independently or aligned with respect to the same period of time (See Fig 7.9).
A Light Touch

Peyton et al. encourage designers of perinatal health technologies to “use immediate needs as a hook for long-term concerns” (Peyton et al., 2014a). This is particularly pertinent advice in the context of design for mental health disclosure and is supported by women’s observations that mental health content and assessments should be ‘coupled’ (P9), ‘couched,’ ‘disguised,’ and ‘sandwiched’ (P11) with other features. Women suggested that a lighter touch was key to inviting honest disclosure, “a tool to sort of mention that people often, kind of are, feel differently during pregnancy and that that’s entirely normal” (P6), “your app needs to be fairly clever about it” (P11).

Promoting Responsibility, Appropriate Expectations & Autonomy

In the clinical context, posing a question often creates a responsibility to act upon the response provided. This requires pathways to support appropriate action and can create an additional workload for midwives as well as tensions and inefficiencies in the shared responsibility of a patient’s care (O’Kane & Mentis, 2012). A mixed group of health professionals took part in our first design session, and discussion often revolved around pragmatic concerns pertaining to clinical practice, including the burden new technologies might introduce; “women, or people rather are generally using apps more and more
in their daily lives so I think it does certainly enhance and benefit them. But we, I think we are, we don't have the, um, ability or access to use apps in our work, never mind the space to get onto the computer” (Female Midwife). One midwife suggested that self-report technologies were most likely to be adopted if responsibility to act was shared across the clinical team;

“If it's a tool to elicit [women's] true feelings, then that's only gonna be good isn't it …but we have to adapt and it would take a bit of planning …for one midwife I think it would be difficult, but if there were a team doing it, that might ameliorate worries.” Female Midwife

The design of this feasibility study was sensitive to the need for technology to be “integrated with …systems people are using” (Female Maternal Health Researcher) and into “existing care, rather than being another thing I have to do” (Male Maternal Health Researcher). Clinicians often hold expectations that communication with patients will provide information which allows them to 'do something' rather than to better ‘understand something,’ an ethos related to concerns about responsibility chains and the capacity to respond in high-risk cases and emergency scenarios. Health professionals therefore attribute significant exigency to data and exhibit a preference for technologies which allow them to use ‘their area of expertise’ and provide a familiar means of interacting with patient data such as blood test results, “they just pull them up on a screen, they get, ehh, this is a score, is it inside or outside a range and they act upon it” (Male Maternal Health Researcher) or routine child protection practices for example, “a red alert or some warning then we know …like with child protection” (Female Midwife).

Another midwife made comments to suggest that supporting midwives’ autonomy would be key to her use of any self-report technology; “if it was something that I had seen someone, and I had asked someone to do for me, then that’s different because I’d then look at the results before seeing them again …so that might be really helpful.” Midwives were also keen to stress that technology not act to disempower patients; “it’s about empowering women to take responsibility for their mood and contacting us, rather than me going ‘oh my god I’ve got to bring in all these women because their mood is very low.’” It is often reported that mental health professionals experience “much more difficulty in interacting with participants without any progress data” (Prochaska & Velicer, 1997) and “difficulty learning from their clinical experiences,” “when they do not receive accurate feedback” or “when their cognitive processes are inadequate (i.e., when they remember information incorrectly)” (Garb, 2000). Access to women's self-reported data might therefore also serve to empower midwives by providing a more accurate picture of their clients' wellbeing. BrightSelf was designed to facilitate actionable information, provide timely alerts, protect patients’ data, and set consistent expectations for women and midwives.
7.2.5 Actively Designing for Engagement

As described in Chapter 4, studies employing self-report technologies often report significant rates of attrition (D. J. Beal & Weiss, 2003; Runyan et al., 2013). This underlines the need to design to engage users, as explored extensively in Chapter 5. Women described the characteristics of various perinatal health technologies which had motivated their use during pregnancy; an appropriate tone, “the wording, their style and the way that they communicate” (P1), broad appeal, convenience and timeliness, a focus on women's needs, and anonymity, “talking about those anxieties ... without risk of insulting your closest friends” (P6). Women were particularly enthusiastic about a feature of several popular apps which compares the size of the fetus each week to a fruit or vegetable (P1, P2, P5, P6, P8). This provides a practical visual analogy, allows women to track their progress, and also enables women to raise the topic of their pregnancy in conversation with others.

Interaction Over Time

Pregnancy implies a longitudinal perspective, and it is therefore essential to consider how interaction unfolds over time. In previous design research however, the suggestion that pregnant women should “actively track and record their activities” was “met with incredulity, laughter and sometimes derision” (Peyton et al., 2014a). Women's perceptions of appropriate reporting patterns varied significantly according to the nature of the scale and their own state of being at the time; “when you're pregnant, you tend to be very tired ... I would be willing to spend like 5 minutes a day to do that but not more” (P4). Several women felt that sporadic reporting 3 times a day for a period of several consecutive days, for example, would often seem too much whereas others believed it might at times prove useful.

In addition to those approaches to the design of engaging experiences described in Section 5.6, women's spontaneous insights often point towards strategies likely to support interaction over time, including positive perceptions of mobile notifications (See Fig 7.10), a suggestion to pair reporting with routine activities such as Kegel exercises, and an impression that reporting could ‘counter-intuitively’ prove more rewarding the more it's used. Developers of perinatal self-report technologies might also
strive to align their design choices with women's perceptions of the temporal character of the antenatal period; displaying the individual week of pregnancy, providing regular updates, and attending to the timeline of perinatal care (See Fig 6.4) as well as women's own narrative experiences (See Fig 6.3).

**Value Beyond Self-Report**

High rates of drop-out and attrition suggest that it is necessary to design for value beyond the practice of self-report, in order to offset the burden of reporting and support a positive user experience (D. J. Beal & Weiss, 2003; Runyan et al., 2013).  

In this instance, our aim was not to maximise the time women spent using an app but to provide value to users and incentivise engagement in the practice of self-report. BrightSelf includes an 'ideas machine,' an animated virtual device which dispenses one of more than 60 'micro-interventions' at the click of a button; tips and brief exercises to support mindfulness, relaxation, and the reframing of negative perspectives (See Fig 7.11). This content was informed by previous patient-centred research (Blake, 2014) and comments from women who described reducing stress as one of their priorities (P2, P8, P9), through meditation (P4) and yoga (P1). This feature of the application was received particularly positively by women, cited as a reason to return to the app, and described as “really really useful” (P11), a “very nice easy pick-me-up” (P4), “very relevant” (P4), ‘fun’ (P5) and “worded in a nice way” (P1).
Midwives were also enthusiastic about this feature although cautioned against including content which made assumptions about or requested too much of women.¹⁰

In addition to providing valuable content for users, this feature serves to lighten the tone of Bright-Self — conceivably, and according to participants’ comments, contributing to a positive user experience, engaging women in self-report, facilitating self-awareness, increasing honest disclosure, and in turn enhancing access to care and support for women in distress while also satisfying the needs and values of health professionals.

7.3 Conclusion | Reflections & Takeaways

The design of engaging tools for the self-report of subjective experience has the potential to enable a more complete understanding of wellbeing, and more effective mental health screening programs; facilitating care and support for those in distress. Working with women, midwives, clinical psychologists, psychiatrists, GPs, and other health professionals, we have illustrated how designers of health and wellbeing technologies might act to support these aims, by attending to the framing of these systems, shaping the experience of self-report, supporting self-awareness and disclosure, providing value to users, and actively designing for engagement over time.

The BrightSelf platform was developed for deployment through NHS midwifery clinics across England, with an eye towards establishing a case for the design and feasibility of mobile applications for the self-report of wellbeing and depression during pregnancy. Chapter 8 presents the results of the first real-world, large-scale and longitudinal clinical feasibility study of this kind.
Chapter 8

Results | The Self-Report of Wellbeing in Pregnancy

This chapter examines the feasibility of a mobile application for the self-report of psychological wellbeing during pregnancy.

8.1 The Study Population

Between April and September 2017, midwives at 14 NHS midwifery clinics across England recruited women to a feasibility study involving BrightSelf. Women attending their first appointment between the 12th and 14th weeks of pregnancy were asked if they would like to use a mobile app for the self-report of wellbeing and depression. Those who consented to take part provided baseline data, including demographic information, and completed both the Whooley Questions and EPDS in-clinic. It was not always possible for women to install the mobile app in-clinic, primarily due to a lack of internet services, and in these cases women were provided a userID with which to install and register the application in their own time.

Of the 355 women who consented to participate, 254 subsequently installed BrightSelf (See Fig 8.1). 128 participants were randomly allocated, in advance, to arm 1 (EPDS reporting only) and 126 to arm 2 (EPDS & EMA reporting).
8.1.1 App Installation by Population

Much of the promise of mobile technologies hinges upon their potential to extend care to at-risk and under-served groups. 71.5% of those women who consented to participate in this study installed BrightSelf. Following the study protocol, we compare the women who provided consent with those who installed the app to examine the relationship between women's demographic characteristics and their tendency to install an app for the self-report of wellbeing in pregnancy.

Women of all ages installed BrightSelf (See Fig. 8.2). There was no evidence of a statistically significant relationship between women's age and their willingness to install a mobile app for the self-report of wellbeing ($\chi^2=9.8109, df=4, p=0.04374, n=348$).
21% of participants reported an ethnicity other than White British (See Fig 8.3). Women of White British ethnicity were more likely to install BrightSelf than women of all other categories combined ($\chi^2=10.58$, $df=1$, $p=0.001143$, $n=348$), possibly due to site-specific or cultural factors.

23% of participants reported 'single' marital status (See Fig 8.4), and there is no evidence of a statistically significant relationship between women's marital status and installation of this mobile app ($\chi^2=2.6289$, $df=3$, $p=0.4524$, $n=348$).
17% of participants reported unemployment (See Fig 8.5), and there is no evidence of a statistically significant relationship between women's employment status and installation of BrightSelf ($\chi^2=10.385$, $df=4$, $p=0.03441$, $n=348$).

![Figure 8.6: Participants by Education](image)

45% of participants reported a level of education below a university or college degree (See Fig 8.6). Again, there is no evidence of a statistically significant relationship between women's level of education and their tendency to install an app for mental health screening during pregnancy ($\chi^2=9.867$, $df=4$, $p=0.04274$, $n=348$).

![Figure 8.7: Participants by Number of Children](image)

45% of those women who installed BrightSelf were experiencing their first pregnancy (See Fig 8.7). Of the 11 women who reported experiencing a previous miscarriage, only 2 did not install the app. There is no evidence of a statistically significant relationship between a woman's number of children and her tendency to install a mobile app for the self-report of wellbeing and depression in pregnancy ($\chi^2=3.6555$, $df=1$, $p=0.05589$, $n=348$).

Similarly, there is no evidence of a statistically significant relationship between app installation and tablet ownership ($\chi^2=0.0878$, $df=1$, $p=0.767$, $n=348$) or a past diagnosis of depression ($\chi^2=3.114$, $df=1$, $p=0.077$, $n=348$).
32 women reported a past diagnosis of depression, 18 of whom installed BrightSelf. Overall, women installed an app for mental health screening during pregnancy regardless of their age, marital status, employment status, education, number of children or past diagnosis of depression.

8.1.2 App Installation by Wellbeing

Participants completed both the Whooley Questions and EPDS at baseline. Whooley question one asks “During the past month, have you often been bothered by feeling down, depressed, or hopeless?”, and question two, “During the past month, have you often been bothered by little interest or pleasure in doing things?” 41 women responded positively to question one (of whom 27 installed BrightSelf) and 40 to question two (of whom 24 installed BrightSelf). There is no evidence of a statistically significant relationship between installation of BrightSelf and women’s responses to Whooley Questions one ($\chi^2=0.38256, df=1, p=0.5362, n=348$) or two ($\chi^2=2.2133, df=1, p=0.1368, n=348$).

At their first midwifery appointment, 22 women scored between 10 and 12 points on the EPDS in-clinic, indicating a possible risk of depression. Fifteen women scored 13 points or above, indicating a probable risk. Six women responded positively to EPDS question ten concerning self-harm ideation. However, the relationship between women’s risk of depression according to the EPDS and installation of BrightSelf was also not statistically significant ($\chi^2=1.7326, df=2, p=0.4205, n=348$). Women in distress were as likely to install a mobile app for the self-report of wellbeing and depression as those with a positive perception of their wellbeing (See Fig. 8.8).
8.1.3 Device Type by Population

Of the 254 women who installed BrightSelf, 68% were iOS device users (32% Android users). This ratio reflects increased rates of iOS device ownership among women within the UK (Statista, 2018). The magnitude of this difference may also highlight the role enthusiasm for technology or another contingent variable plays in women’s choice of device and willingness to install a self-report app. Within the sample of active BrightSelf users, there was no statistically significant relationship between device type and women’s age ($\chi^2=2.6237, df=4, p=0.6226, n=248$), ethnicity ($\chi^2=1.7065, df=1, p=0.1914, n=248$), marital status ($\chi^2=4.9424, df=3, p=0.1761, n=248$), employment status ($\chi^2=4.64, df=4, p=0.3263, n=248$), education ($\chi^2=6.4457, df=4, p=0.1682, n=248$) or tablet ownership ($\chi^2=2.2686, df=1, p=0.132, n=248$).

8.2 Evidence for Feasibility | Women’s Engagement

The primary aim of this study was to examine the feasibility of deploying a mobile application for the self-report of wellbeing in pregnancy. The rich data captured by BrightSelf allows us to triangulate women’s engagement from multiple perspectives.

8.2.1 Evidence for Engagement | Self-Report

Over a 9 month period, women across England made 2,280 reports using BrightSelf; 1,532 reports of their mood, rest, worry, enjoyment and energy in the moment and 748 retrospective EPDS reports. The reporting trend reflects the characteristic context of pregnancy (See Fig. 8.9). The first week of
app use as illustrated here aligns with the 12th to 14th week of pregnancy, and the x-axis spans a period of 9 months. Few studies have examined the contiguous self-report of mental health with respect to such a time-frame in any context. By the 23rd week of app use, many women have reached full-term. A significant number of women therefore engaged with the practice of self-report for a period of many months during pregnancy.

This pattern of interaction demonstrates the value of notifications in bringing women back to an app. The peaks at weeks 5, 9 and 13 are the result of women's overlapping reporting schedules. The potential for overlap decreases over time as the effects of randomisation increase. It is also worth noting however, just how many reports women provided during periods without any notifications, as evidenced by weeks 2 and 3 of app use in particular.
Women in arm 1 synchronised 3.6 reports on average (See Fig 8.10), and women in arm 2, 14.4 (See Fig 8.11). Several women provided more than twice as many reports as they received notifications. Others logged their first report 20 or more weeks after installing the app only to generate an alert for a probable risk of depression.

8.2.2 Evidence for Engagement | Protocol Adherence

Over the course of 6 months, women in arm 1 received 6 notifications. Women in arm 2 received 84. These protocols were chosen to invite and explore different reporting practices. Women were free to adhere to this schedule, provide reports whenever convenient, disable notifications, or indeed to delete the app. During a survey completed post-study, one woman admitted to disabling notifications and providing reports “when I felt like as often I can’t respond when prompted as I’m at work.” This participant’s continued use of an app suggests the importance of supporting users’ autonomy, even in the case of data collection to support research.

Women’s adherence to these protocols also reflects their engagement over time. Figures 8.12 & 8.13 were produced by aligning the notification schedules of all participants and treating as a valid response any report which took place during the period between two notifications of the same type. The spike in EMA reporting at the end of each reporting period is an artefact of this approach to adherence calculation. Each EMA reporting period shown in Fig 8.13 is on average 4 hours in duration. Over a
period of more than 6 months no period elapsed without a response from at least one participant. To facilitate comparisons between arms, we combine the total number of responses within each of the six reporting periods, as shown in Table 8.1.

<table>
<thead>
<tr>
<th>Report Type</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
<th>Period 5</th>
<th>Period 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPDS Arm 1</td>
<td>76</td>
<td>55</td>
<td>42</td>
<td>38</td>
<td>28</td>
<td>24</td>
</tr>
<tr>
<td>EPDS Arm 2</td>
<td>88</td>
<td>41</td>
<td>31</td>
<td>25</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>EMA Arm 2</td>
<td>430</td>
<td>228</td>
<td>185</td>
<td>151</td>
<td>110</td>
<td>86</td>
</tr>
<tr>
<td>Combined Arm 2</td>
<td>518</td>
<td>269</td>
<td>216</td>
<td>176</td>
<td>123</td>
<td>103</td>
</tr>
</tbody>
</table>

Table 8.1: Reporting by Arm

In arm 1, 32% of those who provided a report during the first week of use responded to a final EPDS notification received 15 to 25 weeks later. 19% did so in the case of arm 2. However, women in arm 2 also provided 4 times as many reports as women in arm 1 on average. The addition of momentary reporting led to many more reports on average without greatly reducing the number of EPDS reports provided. These findings appear to mirror women’s comments, made during the design phase, that they would be likely to complete EMA reports more often than the EPDS.

We are able to fit a trend-line, using local regression, to the number of reports provided during each of the six notification periods in each arm (See Fig 8.14). A linear fit, using linear regression, explains 92.4% of the variability around the mean in the case of EPDS reporting in arm 1 ($r^2$ value)), 75.1% in the case of EPDS reporting in arm 2, 82% in the case of EMA reporting in arm 2, and 80.9% in the case of EMA and EPDS reporting combined in arm 2.
8.2.3 Evidence for Engagement | Time Spent

The 'time spent' on app is another possible indicator of user engagement. On aggregate, women spent more than 52 hours interacting with BrightSelf, a total of 2,686 sessions with an average length of 1 minute and 24 seconds. Ancillary usage data allows us to examine exactly how women spent their time (See Fig 8.15). Use of each of the main sections of the app was largely consistent over time, including time spent viewing data, accessing information and interacting with the Ideas Machine.

It should be noted that this application was not designed to maximise ‘time on app.’ Women’s almost ubiquitous use of mobile apps in pregnancy results in an economy of attention in which designers
can feel compelled to compete. As one woman states in the post-study survey, users may have spent
more time on app had we incorporated more additional features, "I often have a flick through my app
‘Baby Centre’ looking at how the baby is developing, reading articles about pregnancy etc.” We chose
however to design for brief, simple and sporadic use based on input from women and professionals. The
average session duration reflects this efficiency of interaction and changed little per week, suggesting a
consistent user experience (See Fig 8.16). In the same post-study survey, 47% of women spontaneously
described quick and easy interaction as the feature of the application which they most liked.

Women synchronised 812 uses of the Ideas Machine throughout pregnancy (See Fig 8.17), and
there is evidence of a significant relationship between the number of reports provided and the use of
this feature (Spearman’s correlation, $S=2140100$, $p<2.2e-16$, $ρ=0.6953086$, $n1=n2=254$).
While correlation does not imply causation, women who used the Ideas Machine more frequently also provided
more reports.

8.2.4 Evidence for Engagement | Population & Wellbeing

Extending public health screening to those in need requires engaging patients regardless of their so-
cial status, cultural background or personal characteristics. In the case of the users of BrightSelf, no
significant differences were found between women’s age ($\chi^2=4.0321$, $df=4$, $p=0.4017$, $n=248$),
etnicity ($\chi^2=3.0946$, $df=1$, $p=0.0786$, $n=248$), marital status ($\chi^2=2.5942$, $df=3$, $p=0.4585$, $n=248$), employment status ($\chi^2=3.8395$, $df=4$, $p=0.4282$, $n=248$), education ($\chi^2=3.9456$, $df=4$, $p=0.4134$, $n=248$), tablet ownership ($\chi^2=3.9456$, $df=4$, $p=0.4134$, $n=248$), tablet ownership ($\chi^2=3.9456$, $df=4$, $p=0.4134$, $n=248$).
Wallis $\chi^2=1.7774$, df=1, $p=0.1825$, $n=248$), number of children (Kruskal-Wallis $\chi^2=0.21977$, df=1, $p=0.6392$, $n=248$) or past diagnosis of depression (Kruskal-Wallis $\chi^2=0.2469$, df=1, $p=0.6193$, $n=248$) and rates of reporting. Women engaged with an application for the self-report of psychological well-being in pregnancy regardless of their demographic characteristics (See Fig. 8.18).

Nor was there a statistically significant relationship between the type of device owned by women (Android or iOS) and their engagement in the form of the total number of reports provided (Kruskal-Wallis $\chi^2=3.1422$, df=1, p-value=0.07629, $n=254$), the total number of Ideas Machine uses (Kruskal-Wallis $\chi^2=3.9258$, df=1, p-value=0.04755, $n=254$), or the total time spent on the app (Kruskal-Wallis $\chi^2=3.1962$, df=1, p-value=0.07381, $n=254$).

Similarly, no statistically significant difference was found between women meeting EPDS thresh-
olds for no, possible and probable risks of depression at 12 to 14 weeks and their engagement in self-report during pregnancy in the case of arm 1 (Kruskal-Wallis $\chi^2=2.3908$, df=2, $p=0.3026$, n=125), arm 2 (Kruskal-Wallis $\chi^2=1.9505$, df=2, $p=0.3771$, n=123) or all users combined (Kruskal-Wallis $\chi^2=0.96303$, df=2, $p=0.6178$, n=248). Nor was there a significant difference between women’s baseline EPDS scores and the total time spent using BrightSelf (Kruskal-Wallis $\chi^2=1.16$, df=2, $p=0.5599$, n=248). Women engaged with a self-report application in pregnancy regardless of their level of ill or wellbeing.

Distributions of women’s use of the Ideas Machine according to their baseline EPDS assessments did differ significantly however (Kruskal-Wallis $\chi^2=9.0126$, df=2, $p=0.01104$, n=248). Women in greater distress were more likely to use this feature, suggesting it was perceived as useful by those in need (See Fig 8.19).

8.2.5 Evidence for Engagement | Users in Distress

Figure 8.20: EPDS Scores by Week | Users Who Provided Multiple Alerts

Of the 748 EPDS reports provided using BrightSelf, 71 met the threshold for possible (41) or probable (27) depression. Two EPDS reports at baseline reflected a risk of self-harm or suicidal ideation. Fifteen did so through BrightSelf. Of the 39 unique women who registered alerts during the study, 7 provided EPDS scores of 10 at baseline, and 6 registered scores of 13 or above. As soon as a woman synchronised an EPDS report indicating a possible or probable risk of depression or self-harm, the research team would receive an email alert. Within 24 hours, and often almost immediately, the appropriate midwifery team was contacted and in turn would establish direct contact with the individual.

26 women therefore received support which they might not have acquired given their baseline
EPDS reports alone. Two-thirds of these women presented no risk of depression according to either the EPDS or Whooley Questions applied independently in-clinic. 56% presented no risk at baseline according to both methods combined. 9% of the women who installed BrightSelf registered EPDS scores of 10 or above on paper at baseline. 16% did so through the use of the app in pregnancy.

Interestingly, of the 39 women who provided EPDS scores which resulted in contact from a midwife, 20 continued to provide EPDS reports and 13 provided at least one subsequent alert (See Fig 8.20). In the post-study survey, women who were contacted by their midwives following an alert spoke positively of the experience;

“I had a call from a midwife to see if I was ok and if I needed any help. I think it should be made available to all pregnant women and not just for research purposes.” Participant No. 294 | Arm 1 | Writing in the Post-Study Survey

8.2.6 Subjective Engagement | Women’s Reflections on the Use of BrightSelf

“I thought this was a brilliant app and an excellent research project. I don’t know what factors have contributed to this but I feel a lot better after having my second child than I did my first.” Participant No. 265 | Arm 2 | Writing in the Post-Study Survey

62 women (37 arm 1, 25 arm 2) responded to a post-study survey seeking their experience of an app for the self-report of psychological wellbeing in pregnancy. The reported experiences of these women provide context for the preceding findings and allow us to rule out contingent effects related to the usability of the app.

92% of women agreed or strongly agreed that the mobile app was easy to use (See Fig 8.21). Similarily, 95% of women agreed or strongly agreed that they learned quickly to use the app (See Fig 8.22). When asked whether the mobile app met their needs 56% of participants responded positively,
with 26% undecided (See Fig 8.23). 65% of respondents specified that they found the assessments useful with 11% undecided (See Fig 8.24).

We also surveyed women’s responses to a number of open-ended questions. Responding to a question regarding their motivations for the use of a self-report app in pregnancy, 22 women described a desire to support research, and 14 to help others; “[I] believe it could be helpful for people and wanted to be part of it.” Women also described personal motivations; monitoring mental health or mood (n=7), supporting self-reflection (n=3), knowing someone who had struggled during pregnancy (n=3), and previous experience of a difficult pregnancy (n=1) as motivations for the use of BrightSelf; “Sometimes I felt low and this help helped me to recognise those feelings.” Several women commented that their motivations changed over time; “Initially it was because I am interested in research and mental health but once I started using the app I felt that it was useful to monitor how tired I felt which reminded me to take time for myself after a busy day at work.”

This post-study survey also provided an opportunity to triangulate women’s qualitative engagement. 53% of woman (with 20% undecided) rated the experience as engaging (See Fig 8.25). 73% (with 10% undecided) reported that they would recommend BrightSelf to a friend (See Fig 8.26), and
69% of respondents (with 15% undecided) stated that they would repeat the experience (See Fig. 8.27).

“At times I was extremely tired where I was working full time and also had to flee a domestic abusive relationship and it seemed inconvenient trying to find the energy and time to complete it but looking back I think it was still very positive.” Participant No. 151 | Arm 1 | Writing in the Post-Study Survey

When asked what they liked most about the experience of the mobile app, 29 women responded that it was quick and easy to use. 14 women stated that the app helped them to engage in mindful reflection; “It made me reflect on the week I had and think about how rational or irrational my emotions had been.” Other responses included the sense that one was making a contribution (n=1), the Ideas Machine (n=2), privacy (n=1), unintrusive interaction (n=2), viewing data (n=4), and monitoring and support (n=4). Women who were contacted by their midwives following an EPDS alert spoke positively of these interactions; “At one point I was particularly down and I then received a call from a professional asking me if I was okay and [sic] they were able to support me with any services,” “When I was having a difficult few weeks, my midwife contacted me and offered extra support.”

Appropriate sampling protocol design is an essential feature of many self-report technologies. As
explored in Chapter 4 however, little prior research has examined users’ preferences with respect to time-based sampling. In this instance, 32% of respondents in arm 2, who received 84 notifications over a 6 month period, reported receiving notifications too frequently or much too frequently (See Fig 8.28). 54% of respondents in arm 1, who received a single prompt once a month, reported that this was too infrequent or much too infrequent a schedule.

When asked what they liked least about the experience of using the app, 25 women referred to the frequency of notifications, although women in both arms described notifications as both too frequent and infrequent. One woman expressed dismay that notifications stopped “towards the end of my pregnancy.” Seven women cited the scales employed, requesting greater variety with respect to both questions and responses. One woman noted that at times of anxiety she felt less eager to use the app, and two requested the ability to reset their passwords. Women’s comments concerning possible improvements to BrightSelf largely mirrored these responses. Nine women referenced different notification frequencies, 8 more varied questions and responses, 8 additional functionality, and 4 clarifying the purpose of the app.

Finally, we asked participants in arm 2 whether they found it useful to compare the reports they
providedinthemomentwiththosemaderetrospectively(intheformoftheEPDS).52%(with20%
undecided)respondedaffirmatively(SeeFig 8.29).

There was no evidence of a statistically significant relationships between women’s actual reporting
rates and subjective ratings of engagement (Kruskal-Wallis $\chi^2=6.4882$, df=6, $p=0.3708$, n=61), willingness
torepeattheexperience(Kruskal-Wallis $\chi^2=8.0513$, df=6, $p=0.2344$, n=61), willingness to
recommend the app to a friend (Kruskal-Wallis $\chi^2=10.682$, df=6, $p=0.09873$, n=61), ability of the app to meet
their needs (Kruskal-Wallis $\chi^2=12.258$, df=6, $p=0.05645$, n=61) or the perception of assessments as
useful (Kruskal-Wallis $\chi^2=2.9494$, df=6, $p=0.8152$, n=61). This may be a consequence of sample size or
selection bias among post-study survey respondents, or simply reflect the diverse patterns of reporting
in which women engaged according to their needs.18

There was a statistically significant relationship between women’s perceptions of the frequency of
notifications and their reporting practice however (Kruskal-Wallis $\chi^2=12.331$, df=4, $p=0.01505$, n=61).
Women who rated notifications as much too frequent and much too infrequent both provided fewer
reports than other users (See Fig 8.30). This finding highlights the importance of appropriate sam-

Figure 8.28: Post-Study Survey Response | Reporting Frequency

Figure 8.29: Post-Study Survey Response | Comparing Momentary & Retrospective Reports

154
pling protocol design, which this combined analysis of notification settings, reporting practice and the subjective experience of self-report provides initial evidence for.

8.3 Conclusion | Reflections & Takeaways

This chapter examines the feasibility of a mobile application for the self-report of psychological wellbeing during pregnancy. Over the course of 9 months, 254 women spent more than 52 hours interacting with BrightSelf, providing 2,280 reports of their wellbeing during pregnancy; 1,532 reports of their mood, rest, worry, enjoyment and energy in the moment and 748 retrospective EPDS assessments.

Women engaged in self-report throughout pregnancy without extrinsic incentives, a contiguous time-frame with respect to which reporting has rarely been studied and which suggests the potential of mobile devices to facilitate the remote screening and assessment of wellbeing and depression throughout the antenatal period. Women engaged with this mobile application regardless of their demographic characteristics or wellbeing. The addition of momentary reporting in daily life led to many more reports on average without drastically reducing the number of EPDS reports provided by women, despite a heavier notification burden. Women in greater distress used the Ideas Machine more often. And, women who used this feature more often provided more reports.

During this study, 71 alerts were generated following a reported risk of depression, self-harm or suicide. 39 unique women received support from a midwife in this way, an experience which they recounted in positive terms. Two-thirds of these women registered no risk of depression at baseline according to the EPDS or Whooley Questions applied independently in-clinic. Responding to a post-study survey, 73% of women (with 10% undecided) reported that they would recommend a mobile application for the self-report of psychological wellbeing to a friend and 69% (with 15% undecided) stated that they would repeat the experience themselves.
This chapter contributes initial evidence for the appropriate design and feasibility of mobile applications to engage women in the self-report of wellbeing and depression during pregnancy, extend care to under-served and at-risk populations, enable longitudinal, momentary and retrospective data collection, overcome stigma, support disclosure, and foster trust between patients and health professionals.
Chapter 9

Discussion & Conclusion

In this final chapter, I summarise the findings of this thesis, describe future work, and reflect on the role of design as a form of critical engagement with technology for the pursuit of wellbeing.

9.1 In Summary

This thesis suggests, for the first time, the feasibility of deploying mobile applications for the self-report of psychological wellbeing and depression in antenatal care. Designing the systems at the centre of this research entailed attention to questions of wellbeing, self-report, engagement, technology, clinical practice and public health. The chapters of this thesis trace an idea from concept to reality.

Chapter 2 attended to the foundations of this work; the conception of wellbeing, its operationalisation in clinical practice, the significance of maternal mental health for wellbeing on a population scale, and the public health approach to bridging the mental health treatment gap. Chapter 3 examined the evidence for stark differences between reports of experience made in the moment, after the fact and ahead of time, and explored how a sensitivity to multiple time-frames might enable a richer understanding of the report, interpretation and curation of experience. This chapter introduced a framework of multiple self-concepts, a practical heuristic and sense-making tool for design which heightens our attention to the importance of how we both live and think about experience and wellbeing. Chapter 4 followed by outlining the design space for self-report technologies, charting previous attempts to implement ecological momentary reporting systems and protocols as a response to problems of reach, access and disclosure. This chapter drew attention to the significant challenges entailed in the design of self-report technologies, including the engagement of users. A concept of interest to researchers and designers of many kinds, Chapter 5 examined the conception, theory, measurement and design of en-
gagement in extensive detail, motivating the adoption of an exhaustive mixed-methods approach to the assessment of longitudinal and voluntary engagement with technology in daily life.

Chapter 6 built upon this knowledge derived from synthesis of the literature by recounting original design research conducted with women, practice and research midwives, medical researchers, clinical studies officers, psychologists, psychiatrists, GPs, and clinicians in maternal and child health, obstetrics and midwifery. The experiences, needs, values and motivations of these participants informed the design of a protocol to support the first clinical deployment of a mobile application for antenatal mental health screening. Chapter 7 turned to the design of the technology at the heart of this research, describing in significant detail the development of a mobile application for the self-report of wellbeing and depression during pregnancy, as well as implications for the design of future health and wellbeing technologies.

Chapter 8 recounted the large-scale and longitudinal clinical deployment of this technology. Over the course of 9 months, 254 women spent more than 52 hours interacting with a mobile application for the self-report of wellbeing and depression during pregnancy, providing 2,280 reports of their subjective experience without extrinsic incentives. These figures reflect the unique potential of mobile devices to engage patients in the self-report of wellbeing and depression, and facilitate public health screening.

Women attending NHS antenatal clinics across England installed and engaged with BrightSelf regardless of their age, marital status, employment status, level of education, number of children, tablet ownership, past diagnosis of depression, or risk of depression according to screening in-clinic. These findings represent initial evidence for the potential of mobile devices to extend care to a significant patient population, including under-served and at-risk groups — a public health priority and essential step towards bridging the mental health treatment gap.

Previous research has suggested that women of ethnic minority backgrounds in the UK may consider professional assistance less appropriate for the treatment of perinatal depression (Shaikh & Kauppi, 2015). In this instance, ethnic minority women were found to be less likely to install a mobile application for the self-report of wellbeing and depression during pregnancy. However, once installed, women engaged with this application regardless of their ethnicity. This interesting finding suggests the need for future design efforts to focus on cultural differences with respect to the design of technologies for health and wellbeing, their introduction into clinical contexts, and the development of a broad portfolio of screening methods.

Prior to this study it was not known whether it was possible to engage women in the self-report of wellbeing and depression during pregnancy, nor for how long. Little previous research has examined
the self-report of mental health for periods longer than several weeks in any context. During this study, however, many women engaged in reporting for periods of 6 months and longer, suggesting the unique potential of mobile devices to support the collection of longitudinal, momentary and retrospective data in daily life. Women’s motivations for doing so often reflected a desire to support both their own wellbeing and that of other women. This study design entailed a mixed-methods approach to the evaluation of engagement, embracing both rigorous quantitative assessment and rich qualitative description of the concept in order to facilitate a detailed portrayal of use and feasibility.

This thesis contributes evidence not just for the feasibility of these systems but for their appropriate design. 47% of post-study survey respondents spontaneously described simple and efficient interaction as the characteristic of this self-report technology which they most liked. Several women commented that they found intrinsic value in the practice of self-report, mirroring motivations expressed by women during the design phase. Women in greater distress used the Ideas Machine feature of the application more often. And, women who used this feature more often provided more reports. 73% of post-study survey respondents (with 10% undecided) reported that they would recommend a mobile application for the self-report of wellbeing in pregnancy to a friend and 69% (with 15% undecided) stated that they would repeat the experience themselves.

Women who rated notifications as much too frequent or infrequent both provided fewer reports than other users. This finding highlights the importance of appropriate sampling protocol design, which this combined analysis of notification settings, reporting practice and the subjective experience of self-report provides some of the first mixed-methods evidence for. The addition of momentary reporting in daily life led to many more reports on average without greatly reducing the number of EPDS reports provided by women, despite a heavier notification burden.

During this study, 39 women disclosed a risk of depression, self-harm or suicide and received immediate midwife support. Two-thirds of participants who received support in this way registered no risk of depression according to standard screening methods employed in-clinic at baseline. These figures suggest the capacity of a mobile application deployed on women’s personal devices to overcome stigma and support disclosure — challenges which many women and midwives consider the primary barriers to care. Through the use of this technology, women at risk of depression, self-harm and suicide were identified during pregnancy at a rate similar to that of postnatal depression diagnosis nationally. This suggests, in compelling terms, the potential advantage of this opportunity to reduce rates of postnatal depression and other serious forms of postnatal mental illness by making support available to women during pregnancy itself.
Those women who received support using this technology spoke positively of the experience and many continued to engage in reporting thereafter. These findings further suggest the potential of mobile devices to efficiently direct care towards those in need at little extra cost while fostering trust between women and midwives;

“Everyone should have access to the app as soon as they find out their [sic] pregnant, great way of communicating, especially for those less inclined to talk to anyone.” Participant No. 349 | Arm 1 | Writing in the Post-Study Survey

Chapter 1 recounted the unique potential of mobile technologies to engage women in the self-report of wellbeing and depression during pregnancy, extend care to under-served and at-risk populations, enable longitudinal, momentary and retrospective data collection, overcome stigma, support disclosure, and foster trust between patients and health professionals. This thesis presents one of the first studies to examine and find evidence for the value of these systems in the real-world context of clinical care. The capacity of these technologies to efficiently direct care towards those in need with minimal midwife input and at little extra cost suggests that, with sufficient preparation, they might be deployed at scale in a relatively straightforward fashion. Some day soon, every woman attending antenatal care might be offered a link to an app.

9.2 Future Work

![Figure 9.1: The BrightSelf Online Interface | The Midwife’s Perspective](image)

This is a complex context for research and design which presents many opportunities for future work.
9.2.1 Qualitative Insight | Midwives’ Reflections

Introducing a new source of patient data into any clinical context entails significant ethical and medical responsibilities. Designers of health and wellbeing technologies must attend to their potential to overburden professionals, dehumanise care, medicalise pregnancy, fail to manage high-risk cases or adversely shape the criteria by which decisions are made. Public health researchers, pregnant women and a variety of health professionals including midwives were involved in the development of the system deployed during this study. Future design research efforts may aim to support the adoption of BrightSelf and similar technologies at scale, including, for example, the conduct of semi-structured interviews with midwives in order to explore their experience of introducing a mobile application to women, and responding to alerts for depression, self-harm and suicide risks.

9.2.2 Future Clinical Studies | Midwife Feedback

This feasibility study suggests the capacity of mobile technologies to engage women in mental health screening during pregnancy with minimal input from midwives. Future configurations of public health systems and services may take many forms however. Clinical interfaces may be required, for example, to support engagement, data management, sense-making, risk-assessment and patient-provider relationships at a national scale. Future feasibility studies might examine health professionals’ experiences with such systems, forms of data and means of data sharing in clinical practice, as well as the provision of remote midwife-led feedback to women’s personal devices through an online interface (See Fig 9.1).
9.2.3 Analysis of the BrightSelf Corpus

This study not only explored new means of public health screening but also produced a corpus of self-reported perinatal well-being and depression of a scale and clinical significance only possible using mobile devices (See Fig 9.2). Complete analysis of this unique body of data, containing more than one hundred thousand data points, is beyond the scope of the thesis and will be reported in future clinical and public health publications.

The Spectrum of Wellbeing in Pregnancy

“well-being science is at a relatively early stage, and we know almost nothing about the upper end of the well-being spectrum.” F. A. Huppert & So, 2013

This is the first use of a mobile application to gather longitudinal data concerning the experience of pregnancy in daily life. Future analysis of this corpus will aim to advance our understanding of the spectrum of wellbeing in pregnancy including which levels of mood, rest, worry, enjoyment and energy might be expected at each week, the kinds of trajectories of wellbeing which present during pregnancy, the points in time at which women are most likely to struggle (opportunities for effective intervention), to what extent variation in wellbeing is influenced by women’s demographic characteristics, the effects of context on momentary reports, and whether the time of day at which assessments are made influences reported wellbeing (possible diurnal patterns).

The Momentary and Retrospective Experience of Depression and Wellbeing

![Graph](image)

Figure 9.3: The Wellbeing of Participant No. 212 by Week | Arm 2

Our knowledge of mental health and wellbeing is inextricably linked to the practice of self-report.
Decades of cognitive psychology research have revealed significant differences between reports made in the moment, after the fact and ahead of time (Shiffman et al., 2008; Ebner-Priemer & Trull, 2009b). Assessing wellbeing during pregnancy therefore necessitates both a longitudinal perspective and an understanding of subjective experience with respect to multiple time-frames (See Fig 9.3). To the best of the author’s knowledge, the BrightSelf corpus is the largest to ever examine wellbeing with respect to both momentary and retrospective time-frames during pregnancy.

An improved understanding of the momentary experience of health and wellbeing is widely believed to possess the potential to revolutionise clinical practice. Chapter 3 presented a framework of multiple self-concepts whose practical and discursive value was employed to support the design of the BrightSelf application and its deployment. This framework will additionally be employed to facilitate the future analysis of this corpus, providing a coherent framework for the interpretation of potentially conflicting reports of experience with respect to multiple time-frames and of the impact introducing momentary assessments might have on retrospective reports of wellbeing — supporting a richer understanding of subjective experience from self-reported data.

Public Health Screening in Pregnancy

“This app is very good and could easily be used by new mums and mums to be.” Participant No. 252 | Arm 1 | Writing in the Post-Study Survey

Future analysis of the BrightSelf corpus will also strive to inform the design of more effective public health screening programs. A broader conception and understanding of ill and wellbeing has the potential to facilitate new opportunities for disclosure, care and research in the public health context (See Appendix C | Wellbeing Data Examples). Analysis of the practice and validity of momentary and retrospective reporting, for example, as shaped by repeated measurement and design, may facilitate the knowledge required to support proactive and evidence-based care.¹

Interaction Metadata and Wellbeing

Finally, this feasibility study also generated significant ancillary data pertaining to the use of a mobile application during pregnancy. This may also be subject to future analysis concerning, for example, links between women’s interaction with a personal device and their wellbeing over a period of many months.
9.3 Speaking to Complexity

This thesis suggests, for the first time, the feasibility of deploying mobile devices in perinatal mental health screening. This statement pertains to the potential of technology at a national scale and I am keen to ensure that it is not accepted uncritically. For all the promise of technology, few among us wish to spend our lives behind a screen. This work is pragmatic, grounded in daily life and attentive to real-world ethical consequences. I hope it is also however, provocative.

Technologies are ‘things’ of a sort which resist definition in all their complexity. On the one hand, actionable, objective, deterministic and causal. On the other, meaningful, subjective, experienced and moral. The tools we create progressively realise medical practice, individual responsibility and normative ethics. The philosopher of science Bruno Latour put it well; technology is society made durable (Latour, 1990). Unlike “a hammer, a locomotive or a chair,” a mobile phone is “composed of writings all the way down” (Latour, 2008). The world is in need of designers who resist reification while condemned to bringing it about.

9.4 Labelling (Ill and Well) Being

![Graphs showing wellbeing data]

Figure 9.4: The Wellbeing of Participant No. 47 by Week | Arm 2

This thesis offers new insight into the experience of wellbeing during pregnancy (See Fig 9.4). The articulation of experience in terms of dots and lines must not preclude, however, awareness of the constructed and culturally contingent nature of wellbeing. While screening for perinatal depression is typically seen in Western culture today as a means of validating women’s experiences and enabling access to care, this label has in the past been associated with practices of patriarchal origins, constraining women to societal expectations of motherhood, denying women the authenticity of a ‘grieving response’ related to multiple losses, powerlessness and status change, and casting aside the socio-cultural and economic
contingencies of childbirth in order to grant medical professionals control over women’s bodies (Shaikh & Kauppi, 2015).

Understandably, women’s reactions to the label of depression can encompass relief, acceptance, questioning and rejection (Lewis, 1995). Designers of wellbeing technologies must attend to how well-being is operationalised in medical practice, understood by patients and made durable by technology. One woman remarked during the design phase that “it's pretty rare that you get presented with the genuine full spectrum of things that you might be feeling, in a way which is compelling.” Showing “that these are normal emotions of pregnancy … that there’s excitement alongside all these other things” she continues, “would be very helpful” (P6). As described in Chapter 2, promoting mental health requires both identifying illness (a pathogenic focus) and promoting wellbeing (a salutogenic focus) (Thieme et al., 2015). We must attend not only to the pragmatic detail of interaction design but to the ways in which labelling and other forms of abstraction shape patient engagement, medical practice and culture.³

9.5 Remaking the Self

Chapter 3 of this thesis drew attention to evidence for the fact that how we see ourselves is continuously subject to change. Pregnancy in particular entails both felt and enacted change. The clinical interfaces of the future will render our selves, in the form of patients, as bits and pixels (See Fig 9.5). These representations permit new forms of knowledge and action, reflect ‘what matters’ to us, and are at times braced between our desire to exert control over our self-presentation and our resistance to the collective imposition of an obligation to do so. These are technologies of the self.³
Describing our health and wellbeing in terms of bits and pixels entails bracketing our felt experience. This is a kind of reductionism which can serve to narrow individuality, divorce us from subjectivity, provide an illusion of control and belie the determinacy of our actions. Women often spoke of a need to know ‘what’s normal’ during pregnancy. We are compelled to turn outwards, seeking a frame of reference by which to make sense of our own experience. However, any representation which frames an individual’s problems as the result of personal inadequacy, or individuality as error, will do little to support disclosure nor bridge the mental health treatment gap.

As datafication shapes a new normal the salience of these implications will fade into the background of our lives. Designers are tasked with remaining aware of the potential consequences of reductionism and acting to avoid dissociation between database and phenomenological selves. We are forever made and remade. The key is to attend, as designers and users, to what we lose and gain.

9.6 Our Orientation in Time

Technology redirects our attention, casting us in and out of the present moment, forward and backward in time. As described in Chapters 2, 3 and 6 in particular, our relationship to the passage of time has implications for our health and wellbeing.

The practice of self-logging often seeks to produce knowledge unavailable to immediate interrogation, through analysis of our patterns of behaviour or felt experience. A permanent record, however, can increase the weight of our past. Health professionals have voiced concerns that digitalisation will threaten professional autonomy (West, Giordano, Van Kleek, & Shadbolt, 2016). Writing a patient’s commentary down changes its significance, its potential, its potency. And with a clearer record of events comes greater scrutiny. A record of our past can constrain our possible futures.

Technology is also often the promise of an idealised future. At the 2015 Digital Health Summit, the healthcare market manager of a major American telecommunications company espoused the advantage of systems which allow you to “see and understand the choices you’re making on a daily basis, which is really who you are: the choices that you make all day long, whether to take the stairs or the elevator, what you will eat or not eat” (Schüll, 2016). Attending to our possible futures can entail the perception of risk in the present, potentially reframing health as a state of perpetual insecurity.

Wellbeing is contingent both upon the time frame of its assessment and in which we predominately reside. Time-scales of weeks, months and years, rather than lives or minutes, have a curious hold on us. To more fully support our wellbeing, design must also attend to our grounding in the present moment.
Pregnant women are less patients in the conventional sense (suffering from illnesses with defined care trajectories) — making midwifery a more open service than many medical specialities. Women and midwives negotiate care.

Midwives, the gatekeepers to many forms of care, employ both formal and informal means of communication; the results of blood tests, and notes scribbled in the margins of women’s records “usually in really bad handwriting, that’s what they do” (P8). Technology can have formalising effects, including the imposition of a more rigid communication structure. We know from ethnographic analysis of emergency medicine, for example, that patients are called upon to justify the legitimacy of their need through appropriate forms of self-presentation, subject to unspoken medical and moral criteria (Hillman, 2014). The introduction of technology will shape these processes, including the ‘correct rules’ of patient behaviour and the patient-typifications on which medical decisions are based (Hillman, 2014).

For health professionals, working under ‘extreme uncertainty’ is often a necessity which must ‘be embraced’ (West et al., 2016). Midwives tend to perceive digital data as more ‘objective’ than information shared in face-to-face interaction. Misinterpretation, with potentially severe consequences for patients’ health, occurs when we fail to recognise uncertainty. The spaces in between the dots and lines of a patient’s chart also tell a story. Making sense of this uncertainty requires knowledge, expertise, and now, digital literacy. Clinicians have a preference for technologies which allow them to use ‘their area of expertise’ and collect information they can ‘act on.’ Midwives do not see themselves as mental health professionals, and increased access to data without the means to interpret it could lead to a crisis of confidence. Studies of clinical decision-making indicate that “confidence is valued over uncertainty” and that there often exists “a prevailing censure against disclosing uncertainty to patients” (Croskerry & Norman, 2008; Kahneman, 2011). These considerations point to the need to design practices around technology, to decide where we can and cannot allow for negotiation in use and what we should and should not fix, while maintaining the salience of the shared goal of a healthy mother and infant.

Health tracking technologies are shifting the management of wellbeing away from clinics and professionals and into the hands of patients (Levina, 2012; Schüll, 2016). This ‘new public health’ is characterised by patient participation. The CEO of wearable technology company Valencell claims that “[o]ld school products show a number but the next generation will simply tell you how you’re doing: You have improved 10 per cent. They’ll give you personalized direction and meaning” (Schüll, 2016). While facilitating self-knowledge, information technologies can also constrain users and instil dependencies
Pregnant women are a particularly vulnerable group whose autonomy is collectivised through public health services and placed in balance with that of (and responsibility for) the fetus. Verbeek notes that technologies around pregnancy, like ultrasound, already facilitate monitoring, creating new kinds of knowledge of the fetus not previously possible (Verbeek, 2008). However, these technologies also take the place of what was once intimately communicated bodily knowledge between mother and baby, even isolating the fetus from the mother. Ultrasound not only constitutes the fetus as a person but as a medical patient, and in some jurisdictions, one with legal rights (Verbeek, 2008). Designers have a responsibility to consider how technology shapes individuals and societies. We are required to navigate tensions of personal and collective wellbeing, while considering that which we act to normalise. Risk or awareness? Autonomy or dependence? Posing these questions does not deny the potential of technology but speaks to the importance of design.

Disclosing mental illness entails vulnerability. As one midwife exclaimed, many women “don’t actually even share some of their mental health issues with their next of kin.” In contrast to many commercial self-tracking technologies, the system described within this thesis was designed for integration within a public health service, to coordinate and improve access to care.9 The practice of self-report in this context can provide opportunities for disclosure, engage patients in conversation, and support decision-making as well as the negotiation of care. Diverse interpretation of perinatal mental illness as a medical condition, psychiatric disorder and social problem shaped by socioeconomic disadvantage, gender inequality and expectations of motherhood precludes reliance upon a single mode of disclosure and requires professionals to “demonstrate flexibility and creativity” in interaction with patients (Shaikh & Kauppi, 2015). Bridging the mental health treatment gap at a population scale requires a diversified portfolio of screening methods, of which mobile devices may well play an important part, but a part.10 A mobile app will not be for everyone — and that’s quite ok.

9.8 The Pursuit of Wellbeing

Today, technologies code aspects of our lives that have never before been quantified; “friendships, interests, casual conversations, information searches, expressions of tastes, emotional responses, and so on” (Van Dijck, 2014). As Koopman writes, “the data that is ceaselessly being produced about us and by us is so important to us in part because we are constantly invited to live our lives through updates, comments, shares, and other manner of informational self-presentation” (Koopman, 2015). This codification of our realities raises important questions as to how we wish to spend our time and what we wish to direct our attention towards.
Many self-tracking technologies entail extending a scientific paradigm to daily life. All research is underpinned by meta-theoretical assumptions and self-tracking is no exception. These technologies embody epistemic values. Graphing the felt experience of populations and individuals implies the propagation of these values at scale, potentially rendering technological values, societal and individual. Computer scientists, those most likely to shape the development of these systems, are most often proponents of logical positivism. This is the pursuit of knowledge through numbers, an epistemology which “concerns itself with time-series data rather than immediate experience; correlation rather than causation; patterns rather than events” (Schüll, 2016).

Many self-tracking efforts aim to uncover how we ‘should’ live through apt description of our lives. Designed to infer ‘ought’ from ‘is,’ these systems defy Hume’s fact-value dichotomy, the claim that ‘matters of fact’ are entirely separate to ‘relations of ideas,’ questions of virtue, right, wrong, good, bad, ill and wellbeing (Hume, 1738). Digitalisation has extended semiotics “to the core of objectivity” (Latour, 2008). To avoid embodying our technologies with an narrow empiricist spirit, we must clearly distinguish the use of a computer science epistemology to seek certain forms of knowledge from the value systems by which we would more fully live.

These tools increasingly raise ethical issues for design. As Calvo & Peters ask “Would we encourage children to track how many hugs they get from their parents? …tracking intimacy will affect how we view and engage in intimacy” (Calvo & Peters, 2014). According to the chief medical officer of United Health Insurance, the challenge with self-tracking technologies is “getting people to use the damn thing” (Schüll, 2016). Vocabularies and technologies of monitoring and tracking can imply and are at times explicitly marketed in terms of our inability to know ourselves phenomenologically or in...
time, dependence upon a larger paternalistic apparatus, or an intrinsic unhappiness with the present moment. Ben Wood, a researcher at market analysis company CCS insight, describes new parents as “low-hanging fruit for gadget-makers because they want to give their babies the best start in life” (Kleinman, 2017). Many fine ethical distinctions are grounded in autonomy. Designers of health and wellbeing technologies must decide how to engage in these new economics of attention and whether to refrain from ‘designing traps’ for users (Madrigal, 2013).

Drawing on the work of this thesis, we also propose a virtue ethics framework for incorporating phronetic knowledge (everyday wisdom) within design research, and design guidelines for navigating tensions of personal information and public health (See Related & Future Publications). This research put situated ‘knowledges’ in conversation with each other, including distinct disciplinary perspectives which each contribute different interests, foci, methodologies, theoretical frameworks and approaches to ethics. Fostering the values which promote human flourishing requires not only establishing which or whose values, but how they are expressed, promoted and even neglected.

Design choices which are ostensibly questions of technology also pertain to lived experience, social practices, normative ethics, and our health and wellbeing. This thesis addresses questions of ends and means, facts and values. In public health technologies we find an example of a design space whose complexity necessitates ethical and epistemological pluralism, clinical integration and medical rigour, paper prototypes and systems fit for deployment at scale, transdisciplinary research conducted in the context of daily life, over realistic time-scales and with voluntarily engaged and representative populations. I argue for a pluralist, conscientious and engaged design practice for the support of wellbeing.

9.9 Conclusion

Maternal mental health has a lasting impact on women’s quality of life, birth outcomes and children’s emotional, cognitive, behavioural and social development. And yet, it is estimated that at least 50% of perinatal depression cases go undiagnosed (Thurgood et al., 2009; NHS Improving Quality, 2015). Suicide remains the leading cause of maternal mortality within the UK (Knight et al., 2016). In a recent survey of women and health professionals, almost all NHS midwives (96%) reported asking women about their mental wellbeing at their first appointment. However, only 1 in 10 women recalled being asked (Boots Family Trust Alliance, 2013).

This thesis examined how technology shapes the self-report of wellbeing, how users engage and are engaged in the honest disclosure of mental health concerns, how healthcare professionals might act
upon reports of psychological wellbeing, and how technology might contribute to our evolving conception of wellbeing and its pursuit. Through collaboration with public health researchers, pregnant women and a variety of health professionals including midwives, this thesis realised the first longitudinal and large-scale clinical deployment of a mobile application for the self-report of psychological wellbeing and depression in antenatal care, contributing initial mixed-methods evidence for the capacity of mobile technologies to engage women in the self-report of wellbeing and depression, extend care to under-served and at-risk populations, enable longitudinal, momentary and retrospective data collection, overcome stigma, support disclosure, and foster trust between patients and health professionals. This thesis resulted in a new understanding of the feasibility and design of mobile technologies for public health screening and research.
Chapter 1 Notes

1. According to the National Study of Health and Wellbeing which assesses the health and wellbeing of the English population every 7 years.

2. Worldwide, the treatment gap for depression is estimated at 56.3%, and for anxiety, 57.5%. However, these figures are believed to represent significant underestimations given the limited data available for many regions (WHO, 2004).

3. Approximately 72% of these costs are attributed to adverse effects experienced by the child.

4. The 'What to Expect When You're Expecting' mobile application has alone been installed more than 15 million times (WhatToExpect, 2017).

5. All other work described within this thesis, including the preparation, conduct and analysis of design sessions, the design and development of BrightSelf, and analysis of the results of the feasibility study was conducted by the author.

6. Numerous academic and practice-based disciplines are implicated in the design of wellbeing technologies. Public health, cognitive psychology and behavioural economics research informs our understanding of wellbeing and mental health assessment. HCI and mental health research elucidates the needs, values and motivations of users and patients. Clinical psychology and midwifery practice grapples with the reality of mental health service provision. And, user-centred design and technology development underlines the successful adoption of new systems and methods.

7. A term coined by complexity theorist Stuart Kauffman to describe the boundaries of possibility within complex systems (Kauffman, 2003).

8. Kuutti describes the history of the user in HCI; a cog in a rational machine in the 1970s, a source of error in the 1980s, a social actor in the 1990s, and, more recently, a consumer (Kuutti, 2001).

9. In Technology as Experience, McCarthy & Wright examine HCI through the lens of a pragmatist philosophy (McCarthy & Wright, 2004), inviting the designer to question whether technology acts to "connect or fragment experience and life," "facilitate unfolding potential, critical perception, and engagement," enriches "our experience of what we already value" or impoverishes it. Rich Coyne's description of pragmatism as "the operative philosophy of the computer world" is also
discussed, designers and developers being "more likely to talk about freedom, community, and engagement (the language of pragmatism) than about formality, hierarchy, and rule (the language of analytic philosophy)."

10. In Addiction by Design, anthropologist Natasha Schüll intentionally adopts what the philosopher of technology Don Ihde calls a "phenomenology of human-technology," through which she aims to avoid "the tendency of strict materialism to treat technology as an autonomous, determining force, while also avoiding the tendency of human-centered approaches to regard technology as a passive, neutral tool" (Schüll, 2012).

11. Including theorists such as Horst Rittel, Melvin Webber (Rittel & Webber, 1973), Bruce Archer (Archer, 1979), Victor Papanek (Papanek, 1972), Herbert Simon (Simon, 1969) and Nigel Cross (Cross, 1999). See (Di Russo, 2016) for a recent history of design thinking.

12. Stefanie Di Russo also makes the case for critical realism as an appropriate foundation for design research in her dissertation Understanding The Behaviour of Design Thinking in Complex Environments (Di Russo, 2016).

13. Designer and filmmaker Hillman Curtis describes design as the process of "making the invisible, visible" (Curtis, 2002).

Chapter 2 Notes

1. As long as pathology was absent there was a tacit assumption among professionals that wellbeing would prevail (F. A. Huppert & So, 2013).

2. It has been argued that the term ‘wellbeing’ has itself allowed psychologists to ‘demedicalise’ the concept of health (Dodge et al., 2012).

3. Eudaimonic — pertaining to ‘the meaningful life’ (Gallagher et al., 2009).

4. Keyes defines social wellbeing as "the appraisal of one's circumstance and functioning in society" (Keyes, 1998).

5. A definition contingent upon cultural and social norms.

6. Measures of wellness tend to comprise factors such as intellectual or spiritual wellness which other interpretations of wellbeing feature less frequently (Cooke et al., 2016).

7. Equilibrium models of wellbeing, including the dynamic equilibrium theory of wellbeing (Headey & Wearing, 1989), the lifespan model of development (Hendry & Kloep, 2002), and the homeostasis of mood and depression (Cummins, 2010), posit an optimal balance between an individual's resources and the challenges they face (Dodge et al., 2012).

8. Hedonic — pertaining to ‘the pleasant life’ (Gallagher et al., 2009).

9. Attempts to align conceptions of happiness and wellbeing have motivated the introduction of terms such as ‘authentic’ and ‘real’ happiness (Shah & Marks, 2004; F. Huppert, 2014). Seligman first defined ‘authentic happiness’ as "peoples' positive evaluations of their lives," including positive emotion, engagement, satisfaction and meaning (Seligman, 2004).
10. There are widespread doubts about the validity of reports of life satisfaction, which typically demonstrate only small to moderate correlation with multidimensional measures (F. A. Huppert & So, 2013). It is not clear how much individuals vary in their interpretation of the best and worst possible life (F. Huppert, 2014) nor whether these measures reflect “well-being, one’s present emotional state, a general personality characteristic such as optimism or extroversion, or some other construct” (Cooke et al., 2016). The weather at the time of report has been shown to influence reports of life satisfaction to the same extent as acquiring a mild disability or doubling income, as has finding a dime, receiving a chocolate bar, and residing in a more comfortable room (to varying degrees) (Feddersen, Metcalfe, & Wooden, 2012).

11. A recent review of wellbeing measures notes that many scales still reflect cultural biases, neglecting the role of family and community, sexual health, socioeconomic, and sociocultural factors, for example (Cooke et al., 2016).

12. Huppert & So adopted an alternative approach to the conception of flourishing, in terms of competence, emotional stability, engagement, meaning, optimism, positive emotion, positive relationships, resilience, self-esteem and vitality, by extracting antonyms of the diagnostic criteria for common mental disorders as described in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) and the International Statistical Classification of Diseases and Related Health Problems (ICD-10) (F. A. Huppert & So, 2013).

13. Personality traits, including optimism, extroversion and self-esteem, have also been shown to demonstrate correlation with measures of wellbeing (Centers for Disease Control and Prevention, 2017).

14. Wellbeing tends to follow a U-shaped distribution according to age; reaching its highest peaks among younger adults and in early old age, demonstrating a possible association with the problem of work-life balance (F. Huppert, 2014).

15. Employment and wellbeing are typically strongly correlated. A positive work environment provides meaning, purpose, challenge, and social relationships (Shah & Marks, 2004; Centers for Disease Control and Prevention, 2017). The impact of the loss of employment on wellbeing ‘far exceeds’ that related to the reduction in income alone (F. Huppert, 2014).

16. The relationship between education and wellbeing is likely mediated by factors such as health, income and social mobility. However, certain studies indicate that an average level of education is linked to higher life satisfaction (Helliwell, 2003; F. Huppert, 2014).

17. Many hope that data gathered through mobile and sensor technologies will provide evidence for how pleasurable, meaningful and engaging different choices and experiences are likely to prove (Schüll, 2016). Economist Richard Layard has argued that knowledge of happiness and wellbeing will enable “the monitoring of trends, the identification of problem groups in the population, and the analysis of why some people are happy and others are not” (Layard, 2010).

18. In 2008, The UK Government Office for Science published the “wellbeing equivalent of five fruit and vegetables a day”; to connect, be active, take notice, keep learning, and to give (The Government Office for Science, 2008).

19. Research suggests that those who possess materialistic values are generally less happy and less satisfied, an effect possibly mediated by the pursuit of extrinsic, rather than intrinsic, goals (F. Huppert, 2014).

20. Dweck & Leggett’s social cognitive theory describes how individuals’ implicit theories shape their goals, and how these
goals determine mastery-oriented or helpless patterns of behaviour (Dweck & Leggett, 1988). Bandura’s theory of self-efficacy posits that “expectations of personal efficacy determine whether coping behavior will be initiated, how much effort will be expended, and how long it will be sustained in the face of obstacles and aversive experiences” (Bandura, 1977). Self Determination Theory (SDT) posits that wellbeing derives from the fulfilment of three basic psychological needs; autonomy, competence and relatedness (Ryan & Deci, 2001).

21. Studies suggest that time spent viewing television, a largely passive activity, has a detrimental effect on life satisfaction (Yang & Oliver, 2010).

22. There is strong evidence that formal training in cognitive behaviour and mindfulness-based cognitive therapies supports wellbeing (F. Huppert, 2014). CBT is primarily used in clinical settings to alleviate the symptoms of common mental disorders, whereas MBSR therapies are increasingly employed outside the clinical context.

23. A 2013 Pew report suggested that nearly 70% of American adults already track weight, diet or exercise while one-third track health indicators such as blood pressure, blood sugar, headaches and sleep patterns (Fox & Duggan, 2013).


25. Diagnosis has been described as the classification of an individual according to “an arbitrary cut point along a continuum,” representing the “severity at which the person is no longer able to function in their everyday life” (F. A. Huppert & So, 2013).

26. Public health services embed these choices within care pathways which determine the shape and experience of care. Strauss et al. in The Social Organization of Medical Work use the term ‘trajectory’ to encapsulate the ‘physiological unfolding’ of a patient’s disease as well as the longitudinal organisation of work in which both patients and professionals participate. Each speciality has its own “characteristic shape” (Strauss et al., 1985).

27. The perinatal period is defined in relation to mental illness as the time between conception and the child’s first birthday (Boots Family Trust Alliance, 2013).

28. The EPDS is a 10-item questionnaire which assesses feelings of guilt, sleep disturbance, anhedonia and suicidal ideation present during the past 7 days (Cox et al., 1987). A score of 10 to 12 points indicates a possible risk of depression. Thirteen points or more suggests a probable risk. Particular attention is paid to question ten which examines self-harm and suicidal thoughts.

Chapter 3 Notes

1. In the terminology of German philosopher Edmund Husserl’s phenomenology, ‘bracketing’ pertains to the perception and representation of events in consciousness; "when we turn attention inward we tend to become phenomenologists, concerned primarily with how we feel rather than what we do” (Hektner, Schmidt, & Csikszentmihalyi, 2007).

2. Art critic Jonathan Crary writes that “[a]ttentive behavior in front of all kinds of screens is increasingly part of a continuous
process of feedback and adjustment within what Foucault calls a 'network of permanent observation'” (Crary, 2001).

3. This has been described as a form of systematic phenomenology (Hektner et al., 2007), capturing “life as it is lived, moment to moment, hour to hour, day to day” (Shiffman et al., 2008) and permitting the “study of the stream of thought or behaviour” (Hormuth, 1986).

4. Design researchers employ a variety of research methods to support the assessment of users’ experiences, including observation, interviewing, directed activities such as storytelling, role playing and collaging, and directed self-report methods, including questionnaires, beeper studies, journalling and think-aloud protocols (Cooper, Reimann, & Cronin, 2007).

5. The designer’s task is then to make sense of these assembled reports of experience. Organisational and interpretive methods are frequently employed, including diagrams, personas, empathy maps, scenarios, storyboards, task analyses, mental models, affordance mapping and use cases (Cooper et al., 2007).

6. The designer finally employs their interpretation of users’ experiences, in the form of implicit mental models and theories, to guide creativity and inform decision-making (Staggers & Norcio, 1993). Methods such as brainstorming, mood boards, prototyping and word association are often employed during the creative process, as can behavioural theory (Consolvo et al., 2008; Hekler et al., 2013) and design guidelines (Consolvo, McDonald, & Landay, 2009) pertinent to the application context.

7. Retrospective reports are often easier to obtain and can facilitate insight into users’ reasoning about their experiences.

8. User experience is formally defined as a “person’s perceptions and responses resulting from the use and/or anticipated use of a product, system or service” [ISO 9241–9210:2010].

9. Snowballing is often employed in the context of narrative review methodologies (Shachak & Reis, 2009).

10. HCI researchers recognise the limitations of human memory and cognition (Chalmers, 2003; Oviatt, 2006). However, the neurophysiology of memory, cognition and emotion is complex, interlinked and not fully understood (Conway & Loveday, 2015).

11. Analysis of time in HCI most often takes place with respect to cognitive absorption (Agarwal & Karahanna, 2000), flow theory (J. M. Pearce & Howard, 2004) and time perception (Nordin et al., 2013). This research therefore typically entails brief time-scales, with notable exceptions such as longitudinal experience sampling (Ebner-Priemer & Trull, 2009b) and models of interaction (Silverstone & Haddon, 1996; Pohlmeier et al., 2009; Karapanos, Zimmerman, Forlizzi, & Martens, 2009; Huang & Stolterman, 2012).

12. Similar framings are also found in dual process theories of cognition such as the Heuristic-Systemic Model of Social Information Processing (HSM) (Todorov, Chaiken, & Henderson, 2002) and the Elaboration-Likelihood Model of Persuasion (ELM) (Petty & Cacioppo, 1986), as well as social cognitive theories (Dweck & Leggett, 1988).

13. The Diagnostic and Statistical Manual of Mental Disorders (DSM) has also faced criticism for depending "largely on clinical judgement” (Ebner-Priemer & Trull, 2009b). Kahneman writes that the feedback psychotherapists receive from patients is "sparse, delayed, or (usually) nonexistent, and in any case too ambiguous to support learning from experience,”
resulting in a 'low validity environment' (Kahneman, 2011). Psychologist Alan Kazdin has remarked that despite "decades of psychotherapy research, we cannot provide an evidence-based explanation for how or why even our most well studied interventions produce change" (Kazdin, 2007).

14. Ecological validity — referring to "the occurrence and distribution of stimulus variables in the natural or customary habitat of an individual" (Hormuth, 1986).

15. The type of analysis facilitated by ecological momentary data, both within and between subjects, has been termed idiographic or ipsative-normative, "where 'ipsative' refers to deviations around the individual mean and 'normative' refers to deviations around the group mean" (Conner, Tennen, Fleeson, & Barrett, 2009).

16. EMA methods have been employed within HCI to support analysis of player engagement (Fischer & Benford, 2009), to help build personalised mobile phone interruption models (Rosenthal, Dey, & Veloso, 2011), to encourage physical activity (Consolvo et al., 2008), to evaluate ubiquitous computing technologies (Consolvo & Walker, 2003), and to inform the design of mobile applications (Sellen & Whittaker, 2010; Peesapati et al., 2010; Isaacs et al., 2013). ContextPhone is one example of an architecture designed to help developers understand users' experiences by "providing context as a resource" (Raento et al., 2005). However, the use of EMA within the design process remains under-explored.

17. Although valid and reliable questionnaires require significant effort to develop.

18. An interview with at least one user of an EMA application would appear to lend support to this use of data. This participant initially expressed disappointment with respect to the system, having hoped it would 'tell' him "how I was feeling, because that's one of the things I struggle with." Later however, he expressed experiencing "a comfortable curiosity in his mood patterns, and 'more confidence in my feelings' " (Morris et al., 2010).


20. During another study, usage of Facebook and Mail apps decreased by more than half when users were asked to report their sessions (Möller et al., 2013). Acknowledging differences between reports of experience made in the moment, after the fact and ahead of time may have the capacity to support behaviour change. However, change itself is most accurately assessed momentarily. As Schwarz writes, "asking patients whether they feel better now than before their treatment is the most efficient way to 'improve' the success rate of medical interventions, which may explain the recent popularity of 'patient reported outcomes' " (Schwarz, 2012).

21. It is important not to overstate the determinacy of this framework. Users' responses are “in most cases - not elicited by the product as such, but by meanings derived from the product" (Desmet et al., 2001). This is a form of 'theory for design' (Zimmerman, Stolterman, & Forlizzi, 2010), and as such, is required to inhabit a rocky middle ground between thick description and precise implications for design. HCI researchers have argued for a richer understanding of users' reported experiences (Elsden, Selby, Durrant, & Kirk, 2016), greater "cross-pollination of ideas and methods" and "the development of better theories" (Hekler et al., 2013).

22. When HCI researchers have studied design with respect to multiple time-frames, reports often pertain to distinct features
of experience and differences within users are rarely studied (Latulipe et al., 2011; Aharony et al., 2011; Van Gennip et al., 2015; Frommel et al., 2015).

23. When we speak of design, often we are referring, in a pragmatic sense, to the design of ‘levers’ for experience (Forlizzi & Ford, 2000).

24. Fig 3.2 incorporates a variety of HCI concepts which might be leveraged to influence the evocation of a particular self. The perception of time itself is another interesting example, given its role in the selection and pursuit of social goals (Carstensen, Isaacowitz, & Charles, 1999). Many technologies shape users’ perception of time, through instant access to information or by modifying the frequency of updates for example (Schüll, 2012). We tend to ‘frame’ our use of technology, even prior to interaction, as ‘mindless or mindful’ in nature (See Jackson, 1998 as cited in McCarthy & Wright, 2004).

Chapter 4 Notes

1. Figures based on National Comorbidity Survey (n=8,098) respondents who sought treatment (Wang et al., 2004).

2. Figures for 9 developed countries (n=22,810). Treatment rates for mental disorders are significantly lower among developing countries — 8.1% in the case of depression (Ormel et al., 2008).

3. FDA guidelines already recommend the use of momentary assessments for patient-reported outcomes (US Department of Health and Human Services, 2006).

4. This potential applies equally to clinical practice and human-centred research. As Miller writes, “How will psychologists be collecting behavioral data in 2025, when more than five billion people are using smartphones …? It would border on scientific malpractice if we were still giving paper-and-pencil questionnaires to a few hundred local college students, recruiting a few dozen people to participate in laboratory tasks, or running Internet studies for people just sitting at desks” (Miller, 2012).

5. Reaching under-served groups supports not just the provision of care but its efficacy, given that “most evidence-based treatments have been developed with samples of patients and providers whose characteristics fail to represent those in routine care settings” (Shoham & Insel, 2011).

6. A recent study of Irish hospital doctors found that 68% would ‘not want others to know’ if they were experiencing mental health problems (B. Hayes, Prihodova, Walsh, Doyle, & Doherty, 2017).

7. Developed according to systematic narrative review. See Section 3.5 for details.

8. Commercial EMA systems include ESM Capture, mEMA, iForm, HarvestYourData, DataField, Qualtrics, SnapSurveys, SurveyAnyplace, SymTrend, movisensXS, MetricWire, Pendragon Forms, LifeData, CS diario, SurveyMonkey, QuickTapSurvey, nestforms, SurveyPocket, LoopSurvey, SurveyGizmo, Opinio, SurveyToGo, SoSci Survey and LimeSurvey. This list is unlikely to prove comprehensive given the range of technologies which might be adapted to support EMA, including messaging services of many forms.
9. The design of EMA systems is shaped by theoretically informed ‘design hypotheses’ (Hekler et al., 2013). Theories of needs and values often refer to intrinsic and extrinsic motivations for use (Botella et al., 2012). Meta-analysis (n=184) has found support for Self Determination Theory (SDT) as a means of studying motivational processes and planning interventions for mental and physical health (Ng et al., 2012).

10. Design values are functional, social, emotional, epistemic, conditional and interpersonal (Fuchsberger, Moser, & Tscheligi, 2012).

11. Unmotivated users may “refuse to participate outright” or “drop out after a few days” whereas motivated participants may also “show greater conscientiousness, agreeableness, or other characteristics that may not make them a representative sample” (Scollon et al., 2009). Values are not just embodied in technology but brought to interaction by users.

12. When first introduced, concerns were raised about the acceptability of these devices. Researchers encouraged participants to refrain from concealing their use of PDAs in order to facilitate use and support the collection of valid data (Trull & Ebner-Priemer, 2013). Problems were rarely reported in this respect however (Koop, 2002).

13. These are questions of compliance, timeliness, context, automaticity, salience and certainty (Bolger et al., 2003; Santangelo et al., 2013) which support a “clearer understanding of the ongoing quality of participation” (Hicks et al., 2010).

14. The link between physical and mental health is described as “ubiquitous across psychiatric disorders” (Ebner-Priemer & Trull, 2009b). Ancillary measures of physical activity have been shown to demonstrate correlation with depression among children and adults (Volkers et al., 2003; Aronen, Simola, & Soininen, 2011; Trull & Ebner-Priemer, 2013; Robbins & Kubiak, 2014).

15. Santangelo et al’s list of desired EMA platform features includes support for varied sampling schemes, multiple alert options (sound, vibration, visual or a combination) and external trigger possibilities (physiological or contextual) (Santangelo et al., 2013). Researchers and users consulted during the design of a mobile app for the collection of sensor and self-reported data (Ohmage) also requested reminder and prompt message customisation, goal setting and feedback functionalities (Ramanathan et al., 2012).

16. Buttons, check boxes, radio buttons, drop-down lists, Likert and visual analogue scales, number wheels and text fields have all been utilised by self-report applications (Santangelo et al., 2013).

17. Half of the users of one patient survey application encountered problems with vertical and horizontal scrolling (Lanzola et al., 2014).

18. The combination of time and event contingent sampling has been described as a ‘layering’ of sampling strategies (Lathia, Rachuri, Mascolo, & Rentfrow, 2013).

19. The most commonly cited behaviour change theories include Ajzen’s Theory of Planned Behaviour (Ajzen, 1991), the Transtheoretical Model of Behavioural Change (Prochaska & Velicer, 1997), the Fogg Behavior Model (FBM) (Fogg, 2009), and the Health Belief Model (HBM) (Janz & Becker, 1984). Other pertinent theories include social cognitive theories, such as Dweck & Leggett’s social cognitive theory (Dweck & Leggett, 1988), Bandura’s theory of self-efficacy
(Bandura, 1977), and dual-process theories of cognition including the Heuristic-Systemic Model of Social Information Processing (HSM) (Todorov et al., 2002) and the Elaboration-Likelihood Model of Persuasion (ELM) (Petty & Cacioppo, 1986).

20. To see how a conceptual model might be employed to support protocol design, I refer the reader to the structure of a stress episode outlined in (Perrez et al., 2000).

21. Not all research employs scales free of ambiguity with respect to their face validity, including, for example, such questions as “Is the momentary situation familiar, typical? Do you feel physically well? Did any specific events occur in the preceding interval? When anticipating the outcome of the following tests, how good will your performance be?” (Fahrenberg et al., 2001).

22. In the case of retrospective reporting, participants typically employ counting strategies for infrequent events and estimation strategies for frequent events, recall events as having occurred more recently than was the case (Stone et al., 1999), and provide more accurate responses when asked to make relative rather than frequency assessments (Scollon et al., 2009) (See Section 3.6.2).

23. During one study of the lived experience of ADHD, participants (n=52) were provided “a log for writing down any erroneous responses that they or their children made” (Whalen et al., 2006). Allowing users to modify or expand upon their data at a later point in time evokes similar design tensions of user engagement, honest disclosure and data validity.

24. The HSM posits motivations for the cognition we engage in (whether heuristic or systemic). These potential sources of biased data include striving to present attitudes consistent with reality (an accuracy motive), to preserve one’s self-concept (a defence motive), and to express attitudes which are socially-acceptable or conducive to one’s social goals (an impression motive) (Todorov et al., 2002).


26. Wilhelm & Perrez describe a series of experiments by McFall, among others, which demonstrated that cigarette consumption increased among students who counted cigarettes smoked but decreased among those who tracked occasions when they resisted smoking (Wilhelm & Perrez, 2013). Hufford et al. introduced a distinction between behavioural and motivational reactivity during a study of problem drinking (n=33) although found no significant effects (Hufford et al., 2002).

27. It is difficult to establish the causality of change without qualitative assessment.

28. Analysis of reporting times can identify diurnal patterns of behaviour such as ‘morningness-eveningness’ effects (Fahrenberg et al., 2001).

29. EMI technologies include PTSD coach (Aguilera & Muench, 2012), SupportBuddy (Aguilera & Muench, 2012), Online Digital Assistance (ODA) (Sorbi et al., 2007), Mobile Therapy (Morris et al., 2010) and Mobilyze! (Burns et al., 2011) among others (Ebner-Priemer & Trull, 2009b; K. E. Heron & Smyth, 2010).
Chapter 5 Notes

1. By explicit definitions, we refer to statements of the form “Engagement is…” or similar.

2. One model of game experience adopts the flow state as a form of utility for the player (Cowley et al., 2008).

3. One study divided participants into ‘highly sensitive’ and ‘less sensitive’ groups based on a personality difference known as electrodermal lability. Participants in the ‘highly sensitive’ group engaged least with strongly expressive agents and most with softly expressive ones (A. Choi, Melo, Woo, & Gratch, 2012).

4. ESM — another term for Ecological Momentary Assessment (EMA) employed more frequently within certain divisions of the literature.

5. A scale developed by Konijn & Hoorn to measure involvement, for example, asks whether a figure “is appealing” and “makes me happy”; a highly context-dependent interpretation (See H. van Vugt et al., 2007).

6. A single publication may examine engagement with respect to multiple application contexts.

7. Within the engagement literature, affect has also been annotated according to three dimensions (valence, arousal and power of control) (Bianchi-Berthouze et al., 2007), defined as a combination of short-term emotion and longer-term mood (J. Hart, Sutcliffe, & De Angeli, 2012), and examined using appraisal theory (Turner, 2010).

8. Each modality was found to significantly encode viewer-engagement. GSR and EEG measures contributed to a similar extent and all three data sources combined resulted in the best performance.

9. While also reducing the potential for reactivity.

10. Studies suggest that users who possess learning goals are more likely to persevere in the face of challenges than those who hold performance goals (Dweck & Leggett, 1988). Positive computing researchers have in turn suggested “helping each user discover what unique conditions motivate him or her” (Calvo & Peters, 2014).

11. Statistical analysis of a variety of metrics pertaining to use and self-report, including longitudinal patterns of momentary and retrospective reporting, protocol adherence, session duration, time spent, usage of various application features and rates of installation, in light of women's demographic characteristics, rates of depression, self-harm and suicide risks, choice of mobile device and more (See Chapter 8).

12. Pre and post-study survey responses, including demographic data, the extent to which women rated their experience of a mobile application as engaging, easy to use, easy to learn and aligned with their needs, described the practice of self-report as useful and timed appropriately, and reported their willingness to recommend a mobile application to a friend and repeat the experience themselves, as well as women's open-ended reflections with respect to their motivations, likes, dislikes and more (See Chapter 8).
Chapter 6 Notes

1. To paraphrase stage designer Es Devlin (Devlin, 2017).

2. This design research was a form of Public and Patient Involvement (PPI) as defined by the Medical Research Council (MRC) and NHS Health Research Authority (HRA). An ethics protocol describing this work was approved by the Head of the Department of Primary Care and Public Health at Imperial College London and the Joint Research Compliance Office (JRCO) coordinator at the same institution. Written, informed consent was obtained from participants prior to each design session. Participants were provided a participant information sheet, given at least 24 hours to decide whether to take part, and their right to withdraw was clearly explained. All data resulting from these sessions was anonymised and stored securely.

3. Kevin Doherty, Marguerite Barry (School of Computer Science and Statistics, Trinity College Dublin), José Marcano-Belisario (School of Public Health, Imperial College London) and Cecily Morrison (Microsoft Research Cambridge).

4. Accessing potentially vulnerable users is widely recognised as a significant obstacle to conducting requirements analysis during the early stages of the development of mental health technologies (Doherty, Coyle, & Matthews, 2010). By engaging a wide sample of health professionals in the design process we were able to obtain phronetic input from those with everyday experiences of vulnerable populations, including younger women, ethnic minorities and women experiencing domestic abuse.

5. Unlike the EPDS, these momentary scales would not be subject to prior validation for clinical use. The lack of validated EMA scales is in general a major obstacle to their use in clinical contexts and this research represents a first step in the direction of establishing their feasibility. A recent systematic review of instruments for the measurement of wellbeing found that few yield reliability coefficients considered adequate for research (.70 or greater) or clinical (.90 or greater) purposes (See Chapter 2) (Cooke et al., 2016).

6. The design of engaging patterns of interaction is an important and understudied component of self-report and wellbeing technologies in general (Karapanos et al., 2009; Möller et al., 2013). There currently exists “a paucity of studies comparing various time-based designs” for ecological momentary assessment protocols and there are “no general conventions,” a fact which Santangelo et al. write is unsurprising, “as the temporal dynamics of emotional and cognitive processes are largely unknown” (Santangelo et al., 2013).

7. The Whooley Questions is a 2-item instrument which screens for depressed mood and anhedonia present during the past month. Respondents are required to answer Yes or No to each question. An affirmative response to either would typically be followed by further assessment, including the use of a validated screening instrument (such as the EPDS or PHQ-9) or referral to a GP or mental health professional (Bosanquet et al., 2015).

8. We would also communicate a participant’s baseline EPDS results to her clinician should she score 10 points or higher, or respond positively to question ten.
Chapter 7 Notes

1. A sentiment often expressed by Gavin Doherty.

2. This online, clinical interface enables midwives to view women’s data and provide feedback directly to women’s devices. It was not employed during this feasibility study and its design will be the focus of future publications.

3. Native applications were developed for both Android (Java) and iOS (Objective-C) in order to support a positive user experience and mitigate selection bias. In June 2017, Android and iOS devices accounted for 96.57% of mobile Operating Systems (OS) in the UK; Android 48.57%, iOS 48% (StatCounter, 2017).

4. This LAMP (Linux, Apache, MySQL and PHP) server is hosted on an Amazon Web Service EC2 instance. Daily backups are automatically created using AWS snapshots. PHP scripting and Amazon’s Simple Email Service (SES) is used to conduct database operations and alert-communication. Secure Shell (SSH) access to the server is highly restricted and protected by private key.

5. Attending to the inscription of “modes of use” which “inhibit or preclude certain actions while inviting or demanding others” (Akrich & Latour, 1992).

6. It must be emphasised that participating midwives represented a variety of ethnic groups, including those discussed, and so their comments reflect both professional and personal community knowledge.

7. The recruiter, often a research midwife, would provide each woman with a 9-digit userID with which to register the app following installation. This block-randomised numerical sequence encodes a unique participant identifier, an identifier for the clinic in which recruitment took place, security features to prevent members of the public from registering the app, and a code for the experimental arm. These codes were generated by the study coordinator in advance of the study and provided to each clinic along with other recruitment materials.

8. Reactivity — whether we are “tapping a phenomenon as it exists, or as it has been transformed by measurement” (Scollon et al., 2009).

9. Design researchers have more recently begun to suggest reframing self-tracking less as a burden and more as “an effortful, yet powerful tool for creative and reflective self-expression, self-exploration, and communication,” (Ayobi, Sonne, Marshall, & Cox, 2018) permitting many different forms of engagement and disengagement; exploratory, productive and counter-productive (W. Smith, Ploderer, Wadley, Webber, & Borland, 2017).

10. Language itself is an essential feature and focus of design. Studies of health and wellbeing have found that users “prefer messages about the benefits of changing (64%)” compared to “the consequences of not changing (34%)” (Muench, Weiss, Kuerbis, & Morgenstern, 2013), favour a sensitive tone with respect to topics such as addiction (Muench et al., 2013), and that more targeted advice can improve outcomes (Kazdin & Blase, 2011).
Chapter 8 Notes

1. Sixteen sites initially agreed to participate in recruitment. Site 6 withdrew at the beginning of the study and site 10, a large trust, never engaged a participant due to a lack of focus on recruitment.

2. This exceeded our initial aim of recruiting 192 active participants, 96 in each arm of the randomised controlled trial. We continued to recruit beyond this threshold to account for the challenges which presented during recruitment, including the installation of the app in-clinic and the time required to develop trust between the research and clinical teams.

3. n=348. Of the 355 baseline data logs collected, 2 lacked participant IDs and 5 featured empty data fields, likely due to a combination of poor internet connections in-clinic, bugs or crashes in the proprietary survey software, and errors made by recruiters.

4. n=248. Of the 254 unique users who installed the app, 6 (3 arm 1, 3 arm 2) could not be linked to their demographic data provided at baseline, as described above. For consistency, we exclude the data of these 6 women from this analysis.

5. We employ a chi-square test for independence given nominal dependent and independent variables. To protect from Type I errors, we conduct Bonferroni correction for multiple tests (8 tests on the single dependent variable of ‘app installation’) which yields a corrected alpha-value of 0.00625 for statistical significance in this case. While not all tests yielded statistically significant results according to this threshold, this analysis also serves to highlight population characteristics for additional future consideration and analysis.

6. It should be noted that these figures are intrinsically conservative as this dataset contains only that data which users synchronised with the research server.

7. n=254. Here, the first 8 days of use are represented separately as ‘Week 1.’ This is for two reasons; 1) The first day of use effectively represents a significant outlier given that all participants necessarily interact with the app on the day it is installed, and 2) During the first 8 days of use, the notification schedules of all participants are aligned. Subsequent notification periods are subject to randomisation by week within and between-subjects (a possible variance of 3 to 5 weeks at the second notification period increases to 10 to 15 by the sixth). This also has the effect of exaggerating the decline in use over time past the first week. Every subsequent week as graphed represents a period of 7 consecutive days since the morning of the day of installation.

8. These calculations include those women who synchronised no reports.

9. Arm1, n=120, Arm2, n=117. Unexpected behaviour in iOS10 led the data of 17 women (8 arm 1, 9 arm 2) to be returned in 12 rather than 24-hour time format, rendering their timestamped data unreliable to the hour. For the purposes of consistency we omit these users’ data from this analysis.

10. This calculation excludes 188 outliers, interactions shorter than 5 seconds in duration.

11. Spearman’s rank-order correlation is the non-parametric version of the Pearson product-moment correlation, employed in the case of non-normal continuous dependent and independent variables.
12. Reporting rates, total time spent on app and use of the Ideas Machine data do not follow normal distributions (as confirmed using quantile-quantile plots and the Shapiro-Wilk test). We therefore employ a Kruskal-Wallis test for independence as appropriate in the case of continuous dependent and ordinal independent variables with more than two groups.

13. This calculation excludes 13 outliers, instances during which women spent more than 10 minutes on a single screen of the app.

14. To protect from Type I errors, we conduct Bonferroni correction for multiple tests (9 tests on the single dependent variable of ‘app engagement’) which yields a corrected alpha-value of 0.0056 for statistical significance in this case.

15. The survey responses of the two participants who did not rate BrightSelf as easy to use appear to contain inconsistencies. Both specified spontaneously in subsequent open-ended response that the mobile app was “Quick and easy to use.”

16. Again, the single participant who ‘strongly disagreed’ in this instance rated the app positively in her open-ended comments, stating that there was ‘nothing’ she disliked about the experience. This participant provided inconsistently low ratings to all subsequent responses. The other user who provided a negative response in this instance, a participant in arm 1, explained that she felt notifications were sent ‘much too infrequently’ and as such wasn’t sure when to provide reports.

17. Seventy-one motivations in total were expressed by 61 women.

18. Due to a bug in the third-party survey software, a single user’s response lacked the userID field, which did not allow us to compare their reflections to their use of the app.

Chapter 9 Notes

1. This data will also enable analysis of the validity and efficacy of different screening methods, including changes in the internal consistency of EMA and EPDS scales, the relationship between EMA and EPDS scores, the time spent completing EMA and EPDS reports, the number of times a question was revisited while completing EMA and EPDS reports, and delays responding to EMA and EPDS notifications over a 9 month period. We will also assess the relationship between the Whooley Questions and EPDS as completed by women at baseline, including positive predictive values, negative predictive values and false omission rates.

2. The extensive use of exclamation marks in the realm of wellbeing applications (‘Looking good today!’) often precisely fails to speak to the fullness of experience.

3. French philosopher Michel Foucault described ‘technologies of the self’ as those which permit “individuals to effect by their own means or with the help of others a certain number of operations on their own bodies and souls, thoughts, conduct, and way of being, so as to transform themselves in order to attain a certain state of happiness, purity, wisdom, perfection, or immortality” (Foucault, 1988).

4. Cultural anthropologist Natasha Dow Schüll writes that wearable trackers invite us “to view ourselves as longitudinal databases constantly accruing new content: ‘You are your data’” (Schüll, 2016). According to French philosopher Gilles Deleuze we risk reducing the self to a set of endlessly divisible data points, not individuals but “dividuals and masses,
samples, data, markets, or banks” (Deleuze, 1992).

5. Systematic comparison to the (perhaps hypothetical) mean can risk conceiving of mental health and illness in terms of deviance from socio-cultural norms, conflating the ‘average’ mental health of a population (which includes the presence of psychopathology) with health, confronting the unacceptable notion of average in a competitive society, and ultimately proving counter-productive, given evidence of a negative relationship between social comparison and happiness (Calvo & Peters, 2014).

6. Socio-technological critique often pertains to questionable assumptions of modernity, including mastery. Floridi writes that modernity offers two contradictory accounts of the self; “On the one hand, in the political realm, the self is deemed to be free, and ‘free’ is frequently understood as being autonomous, disembodied, rational, well-informed and disconnected: an individual and atomistic self. On the other hand, in scientific terms, the self is an object of enquiry among others and, in this respect, is deemed to be fully analysable and predictable” (Floridi, 2015). Sloterdijk’s envelope metaphor avoids essentialising technology by refusing to cast its action in mechanical terms of determinism (Latour, 2008).

7. Anthropologists and sociologists draw attention to the means by which technology as benign as a ‘smart fork’ can turn “something as routine as a single bite of food into a matter of potential danger” (Schüll, 2016), lead to a ‘regime of anticipation’ (V. Adams, Murphy, & Clarke, 2009), create the sense that “some, perhaps all, persons, though existentially healthy are actually asymptotically or pre-symptomatically ill” (Rose, 2009), and inflict a “double insecurity” of “always being at risk” while “never knowing enough about what one could and should be doing” (Dumit, 2012).

8. HCI researcher Liam Bannon calls for designers to consider the value of forgetting, writing that “[p]erhaps all social messages may be designed to fade away over time” out of a shared concern that “the variety of human interaction means, from informal remarks in a café or bar, to formal memos in an office, might paradoxically be reduced rather than enhanced through technological means” (Bannon, 2006).

9. As populations grow older, more overweight and chronically ill, pressure on services is increasing. The coordination of care was the most sought-after role for technology in a recent survey of Swedish healthcare personnel (718 nurses, 57 midwives, 301 doctors) (PricewaterhouseCoopers, 2016).

10. A diversified portfolio of screening methods may include many means of disclosure including face-to-face conversations, paper questionnaires, and phone and online supports.
Bibliography


193


Matthews, M., Voida, S., Abdullah, S., Doherty, G., Choudhury, T., Im, S., & Gay, G. (2015, August). In Situ Design for Mental Illness: Considering the Pathology of Bipolar Disorder in mHealth Design. In


Stiubiener, I., Silveira, R. M., Matushima, R., Bressan, G., & Ruggiero, W. V. (2012, October). Evaluating the Effectiveness of the Use of Teaching Materials in Video Format in Distance Learning En-


A | Design Session Structure

- **Framing the Design Challenges (5 mins)** The researchers began each session by introducing themselves, the purpose of the session, and several examples of the challenges pertaining to the design of technology in the perinatal context;
  - C1: To maintain the engagement of both women and clinicians
  - C2: To achieve a balance between the burden and utility of interaction as perceived by users (e.g. a balance between clinicians’ need for data derived from a validated scale to inform clinical decision-making, and the burden on women to report data which might appear of little intrinsic value)
  - C3: To collect quantitative data without reducing women and clinicians to their digital roles and representations

- **Starting the Conversation (15 mins)** Participants were then provided the opportunity to introduce themselves and share their interest in the topic.

- **Group Discussion & Identification of Themes (20 mins)** Participants were next divided into groups of three, and supplied with post-its on which to keep note of ideas and key concepts. Discussion revolved around four themes;
  - Care During Pregnancy
    - I1: To understand what clinicians perceive as their primary goals and responsibilities
    - I2: To understand the nature of the relationships between women and clinicians
    - I3: To understand the critical success factors for embedding new systems into existing healthcare practices
  - Motivation & Adoption
    - I4: To understand the main wellbeing concerns of parents, as viewed by women and clinicians
• I5: To understand women’s motivations for the collection of data concerning well-being and depression during pregnancy

• Self-Report & Psychological Wellbeing

  • I6: To determine how both sets of users might see mobile reporting fitting within their daily lives, schedules and care pathways

  • I7: To explore which wellbeing-related concepts women and clinicians consider worth reporting and why

  • I8: To grasp the perceived value of momentary (compared to retrospective) reports to individual women and for clinical practice (including discussion of reactivity, context, time, and other determinants of validity)

• Implementation Specific Concerns

  • I9: To investigate which features of perinatal technologies users consider most valuable (e.g. social features, data visualisation, feedback)

  • I10: To determine which forms of automated and midwife-led feedback users perceive as valuable

  • I11: To examine concerns of look, feel, structure, content, functionality, inputs, outputs, and interactivity with respect to design

• Visual Ideation (15 mins) Blank paper templates and a selection of sketches of the mobile application were supplied to each group, the motivations behind the technology were briefly introduced, and participants were invited to discuss these concepts, annotate the sketches provided and present ideas of their own.

• Prototype Evaluation (10 mins) The current mobile application prototype was presented, on a large screen where possible, and on devices shared among the group. Participants were encouraged to provide their thoughts in a think-aloud fashion while interacting with the system.

• Final Comments (5 mins) Finally, participants were thanked for their time and granted the opportunity to provide any final comments.
B | Design Work

The following pages of Appendix B present images, screen flows and design rationales pertaining to the BrightSelf mobile application.
"when you say the word ‘mental’ or ‘mental health services,’ or anything with the word ‘mental,’ it has a negative sound to me" (P8)
BrightSelf

Dosis, a friendly, rounded font was chosen to mitigate medical perceptions of the app
A dual weight typeface accentuates the twin meaning in the logo

Splashscreen

Apple guidelines do not recommend the use of splashscreens, and yet many popular apps maintain their use
To give the app a professional, familiar feel, and to suggest that the user may find value in the use of the app beyond the collection of data to inform research, we chose to include one

Sun Rays

Bright rays represent hope and positivity

Colour Scheme

A complementary colour scheme of greens, orange, red and blue was chosen to invite trust and engagement while avoiding a stereotypical pink and flowery styling (See Peyton et al., 2014)

A Moon

The moon represents the opposite half of the emotional spectrum; night as to day

A Baby Sling

The constituent shapes of the logo form a Gestalt representation of a baby in a sling
The appearance of the logo is animated on app launch, a flow designed to suggest the meaning behind each component part
A 4-digit pin provides a sense of security
Large numbers allow for quick, single-handed and responsive entry for distracted users

Profile Icon
A flat icon adds some aesthetic interest to the screen

Username
The user’s first name is included to foster trust and suggest a personal, private space

PinCode
A 4-digit pin provides a sense of security
Large numbers allow for quick, single-handed and responsive entry for distracted users
Scenic photographs change each time the app is relaunched.

A changing background image provides a degree of novelty which may support engagement.

ART therapy suggests that even computer-based images of nature can improve mood.

Initial prototypes included ‘inspirational’ quotes here.

Several midwives deemed this patronising and it was removed.

Feedback Version

A third version of the app presents a scrolling feed on this screen which can display feedback from midwives.

Navigation Icons

Icons were designed to convey meaning at a small scale.

Two solid colour circular icons provide symmetry and distinguish the ‘Home’ and ‘Extras’ sections.

Syncing Status

Icon reflects the now common association of ‘cloud storage’.

Placed prominently to remind users of the need to sync data, and also to imply the priority of this action.

Text confirms the meaning of the icon.

Status Bar

A transparent status bar was chosen to support users’ ‘presence’ within the app.

Text confirms the meaning of the icon.

Single Navigation Bar

A one-handed and distracted user was assumed.

This enforced a minimal and consistent layout to support efficient navigation.

The current tab is highlighted to convey the user’s location within the app.

An orange horizontal highlight distinguishes the navigation bar from the background.

HOME SCREEN
Background
A plain white background maintains the user's focus on the purpose of this screen. Significant whitespace provides an impression of elegance and efficacy.

Subtitle
A subtitle in question form further conveys the purpose of this section of the app and draws the eye to the button above.

CheckIn Button
A simple, centrally-located button provides immediate feedback to the user about the purpose of the screen and the action to take. A drop-shadow invites interaction. An uppercase typeface further distinguishes this interactive component from those which explain its use.

Detail Text
Smaller text provides an additional rationale for the use of this mode of report, striking an appropriate tone.

CheckIn Icon
A location pin emphasises the pinpoint nature of this form of report in time.
Instructional Text
Brief text indicates how long these scales will take to complete, important for the engagement of users, and demonstrating respect for their time.

“I think at the beginning stating how long people can expect to do it [to complete a scale] is useful” (Female Midwife)

Background
A solid colour background distinguishes this screen from the previous, and demarcates this separate section of the app.

Reflective Question
A first-person phrasing indicates the personal nature of this reflection and hints at the value in doing so for the individual.

Bold text emphasises the present nature of these assessments.

“I think ’I’ is better than ’You’ because it makes the person think that they’re actually talking to themselves as opposed to talking to an external thing ... I would be more honest with ’I’ than ’You’” (P9)

Previous Button
The user is given the option to leave this section straight away, respecting their autonomy, one of three primary psychological needs according to Self Determination Theory (See Ryan & Deci, 2008).

Next Button
A slightly larger arrow icon suggests the way forward.
The circular outline invites interaction, a theme maintained throughout the app.
Scale Icons
Flat icons provide a graphical representation of each point in the scale. Each Likert scale employs a unique set of icons. Icons are animated, increasing in size when the slider is changed and bouncing back when released, supporting engaging and intuitive interaction.

Previous Button
Tracking revisited questions can provide insight into users’ patterns of reflection, potential usability concerns and reactivity to reporting.

Background
A simple white background supports efficient interaction, and mitigates distraction.

EMA SCREENS

Question Text
First-person oriented text guides the user’s response and promotes honest disclosure. The primary signifier of meaning for each question is highlighted in green to support more efficient interaction.

Anchor Points
Provide additional feedback regarding the valence of the scale.

Scale Labels
Text labels also indicate the meaning of each scale point.

Slider Track
An aesthetically minimal scale supports efficient and single-handed interaction. A change in colour based on the position of the thumb further indicates scale valence.

Next Button
A slightly larger icon encourages the user to progress to the next question.

EMA Questions
5 Visual Analogue Scales (Mood, Worry, Rest, Energy, Enjoyment)
2 Radio Button Scales (Semantic Location, Activity)
An animated horizontal bar chart presents the user with a visualisation of their results and provides a sense of accomplishment.

Simple feedback chosen to support engagement and self-insight.

A minimalist aesthetic is maintained.

Valenced anchor points were chosen over numbers to better reflect a continuum of wellbeing.

Immediate feedback provides value and incentive while respecting each user’s individual circumstances.

More detailed text suggests possible actions in the case of highly negative results, and promotes self-awareness otherwise.

This text is semi-randomly chosen from a set of multiple possible responses based on a labelled sum of user’s EMA reports.

A single line of feedback in first-person provides context for these results.

This text is semi-randomly chosen from a set of multiple possible responses based on a labelled sum of user’s EMA reports, serving as an incentive for repeated use.

A link to see more results hints at the value of maintaining the practice of self-report over time.

A link to see more results hints at the value of maintaining the practice of self-report over time.
CheckBack Screen

Background
A plain white background maintains the user’s focus on the singular purpose of this screen.

Subtitle
A subtitle in question form further conveys the purpose of this section of the app.

CheckBack Button
A simple, centrally-located button provides immediate feedback to the user as to the purpose of the screen and the action to take. A simple drop-shadow invites interaction. A vibrant red theme distinguishes this section of the app from the Check In section.

Detail Text
Smaller text provides a further rationale for the use of this mode of report.

CheckBack Icon
A calendar icon with the number 7 emphasises the retrospective nature of this form of report. CheckIn and CheckBack features are located side-by-side to encourage comparison and therefore understanding.
A solid colour background distinguishes this screen from the previous, and demarcates this separate section of the app.

A vibrant red theme distinguishes this section of the app from the EMA section.

**Instructional Text**
Brief text indicates how long this scale will take to complete.

**Previous Button**
The user is given the option to leave this section straight away, respecting their autonomy and providing a sense of control and security.

**Next Button**
A slightly larger arrow icon suggests the way forward.
The circular outline invites interaction, a theme maintained throughout the app.

**Reflective Question**
A first-person phrasing indicates the personal nature of this reflection and hints at the value in doing so for the individual.

Bold text emphasises the retrospective nature of these assessments.
EPDS SCREENS

EPDS Questions

A 10-item scale for perinatal depression screening

Question

Red text draws the eye to each question

Navigation

Users can go back and forth between questions

Answer

Radio buttons allow for efficient interaction
The use of circular elements to invite interaction is continued
No response is highlighted when the scale is first presented, to mitigate bias
An animated visualisation shows the user’s response to each question, overlayed upon an aggregate score. The choice of whether to display numerical results and their meaning featured strongly in discussion with women and midwives. CheckIn results are animated horizontally, CheckBack results vertically, further distinguishing both.

**Anchor Points**
Relative (interpretivist) rather than numerical (positivist) scoring.

**Text Feedback**
A brief non-stigmatised description of the results, mitigating over-zealous interpretation.

**Detail Text**
Additional feedback chosen semi-randomly based upon the user’s EPDS score.
Ideas Machine

A crate emblazoned with a lightbulb icon, when clicked expands through a colourful animation into an eccentric device which spouts a short section of randomly coloured text which then floats slowly up the screen.

The mechanism of interaction (click-based, semi-random, and delayed through animation) was designed to support engagement and invite exploration.

Additional Features

The purpose of this section of the app was to provide additional value to the user.

Over 60 simple tips, ‘micro-interventions,’ most based around mindfulness tips and activities were included and revealed in colourful fashion (See Blake, 2014).

Idea Text

One of more than 60 tips and activities chosen to endorse self-awareness, self-care, and support-seeking.

This semi-random and numerous selection supports continued use.

Light-hearted presentation attempts to avoid a medicalised or obsessive focus.
Extras

Supplementary features designed to engender trust, inspire engagement, and support users. Separate presentation of these components denotes them as separate to the core content of the app.

Icons

Simple icons on a white background. A text subtitle supports meaning. A circular form prompts interaction.

My Data

A larger icon indicates the relative significance of this section. An orange colour scheme is used to denote all ‘extras’ components.
The motivations behind this research, its funding sources and those involved in it may support engagement. Establishing users’ trust, a working alliance, and emphasising users’ contribution to science may support engagement. Images of the researchers were included in an attempt to establish trust.
Opening Image
Aesthetically engaging

Support
Links to several NHS and mental health charity support options, each with images, web links and phone numbers

"So even the links for the NHS can be just provided" (Female Midwife)

More
Scrolling vertically displays additional content
BrightSelf

Information about the app, its content and features
This section guides, informs and sets expectations for the user, and may also be used to introduce the app to others.

Opening Image
Aesthetically engaging and maintains the user’s interest more effectively than text alone.

Background
A diagonal gradient rather than whitespace signals to the user that this section of the app requires a slower form of engagement.
Considering, and supporting, your mental health during pregnancy is very important. However, the experience of many emotions; fear, anger, sadness, tiredness, worry, doubt, and stress is normal at this time, regardless of circumstances. Even while using BrightSelf, it can be useful not to reflect too much upon any normal changes in your mood. However, in some cases these emotions can be too overwhelming and result in perinatal depression.

"It’s pretty rare that you get presented with the genuine full spectrum of things that you might be feeling, in a way which is kind of compelling”

(P6)
My Data
The user’s wellbeing visualised over time

Graph
An animated line graph showing individual reports and trends over time

Extra Info
A popup provides an explanation of the chart and the data it draws on

Day v Month
Data display by day or month
The user may choose the time-frame for which to display their data

CheckIn Display
The choice of CheckIn concept to display

CheckBack Display
The option to display EPDS results
Displayed as a red bar super-imposed upon the 7-day period to which they pertain
The CheckBack plot can be displayed directly behind the CheckIn plot to facilitate comparison

Interaction
A simple symmetrical on/off interaction mechanism was chosen to facilitate the display and interpretation of this data
The colours of each button match those of the data displayed

My Data
The user’s wellbeing visualised over time
C | Wellbeing Data Examples

The following pages of Appendix C present examples of the data provided by women using the BrightSelf mobile application during pregnancy.
User 11 | Arm 2 | iPhone
49.52 Minutes | 38 EMA Reports | 7 EPDS Reports | 13 Ideas Machine Uses | 0 Alerts

The Wellbeing of Participant No. 11 | Arm 2
The Wellbeing of Participant No. 151 | Arm 1